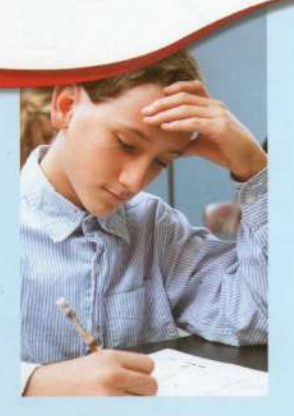
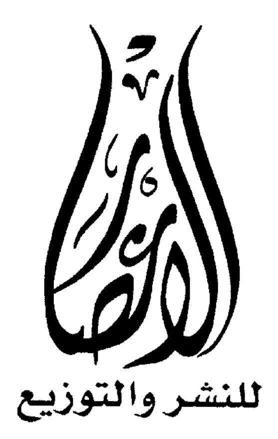
بناء الاختبارات والمقاييس النفسية والتربوية وتقنياتها



الأستاذ الدكتور
امطانيوس نايف مخائيل
أستاذ القياس النفسي والتربوي في جامعة دمشق - سوريا
وجامعة جدارا - الأردن




للنشر والتوزيع


MOHAMED KHATAB





بناء الاختبارات والمقاييس النفسية والتربوية وتقنيتها


MOHAMED KHATAB

بناء الاختبارات والمقاييس النفسية والتربوية وتقنينها

تأليف

الأستاذ الدكتور

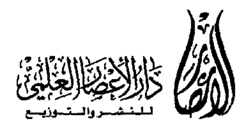
امطانيوس نايف مخائيل

أستاذ القياس النفسي والتربوي في جامعة دمشق –سوريا

وجامعة جدارا –الأردن

الطبعة الأولى

2016م – 1437هـ



رقم الإيداع لدى دائرة المكتبة الوطنية (2014/9/4564)

370.15

مخائيل، إمطانيوس نايف

بناء الأختبارات والمقاييس النفسية والتربوية وتقنيتها / إمطانيوس نايف مخائيل.- عمان: دار الإعصار العلمي للنشر والتوزيع، 2014



الطبعة العربية الأولى

2016م- 1437هـ



الأردن – عمان – وسط البلد – شارع الملك حسين – مجمع الفحيص التجاري

هاتف: 96264646208+ فاكس: 96264646470+

الأردن – عمان– مرج الحمام – شارع الكنيسة – مقابل كلية القدس

هاتف: 96265713906+ فاكس: 96265713907+

جوال: 797896091 - 00962

info@al-esar.com – www.al-esar.com

دار الاعصار العلمي

ISBN 978-9957-98-072-6 (ردمك)

المحتويات

الموضوع	الصفحة

الموضوع	الصفحة

الفصل الثالث

بناء الاستبانة وتوفير مستلزمات صلاحها

الموضوع	الصفحة

الفصل الرابع

بناء اختبارات التحصيل

الفصل الخامس

الصدق

الفصل السادس

الثبات

الفصل السابع

معايير الاختبارات النفسية

الموضوع	الصفحة

الفصل الثامن

استخدام الإحصاء في تحليل نتائج الاختبارات وتفسيرها

مقدمة

ما من شك في أن الاختبارات النفسية تمثّل نقطة تحول مهمة في تاريخ تطوّر علم النفس بوصفها الوليد الطبيعي للنهج القياسي الكمي الذي اختطه هذا العلم. وقد أعطت العقود الثلاثة الأخيرة من القرن الماضي والسنوات الأولى من هذا القرن دفعاً جديداً وقوياً لحركة القياس النفسي، وأظهرت انتشاراً واسعاً للاختبارات النفسية قلما شهده غيرها من المجالات العلمية الأخرى وهذا ما أسهم في بلورة مفاهيمها الأساسية ورسوخ المرتكزات النظرية التي استندت إليها، كما جعلها تحقق تطورات هائلة ومتسارعة سواء في أساليب تصميمها وبنائها، أم في التعديلات والتحسينات التي أدخلت عليها بصورة متتابعة من خلال الدراسات السيكومترية التي أخضعت لها، وما زالت، تخضع لها. ويبدو واضحاً من هذا المنظور أن دراسة المفاهيم والمبادئ التي ترتكز عليها حركة الاختبارات النفسية، وتتبّع هذه الحركة في مساراتها الحديثة والمتقدّمة، ولاسيما من خلال الوقوف عند طرائق تصميمها وبنائها، والإحاطة بالتعديلات والتحسينات المتلاحقة التي باتت السمة المميزة لها — أمر ينطوي على قدر كبير من الأهمية. ولعل مما يزيد أهمية مثل هذه الدراسة أن المكتبة العربية ما زالت تفتقر إلى أدبيات ومراجع القياس النفسي، بما فيها تلك التي تتناول مسائل تصميم وبناء الاختبارات النفسية بصورة ملفتة للنظر.

يتضمن الكتاب الحالي ثمانية فصول. يتناول الأول منها الاختبارات النفسية من حيث خلفيتها التاريخية والعوامل والظروف التي أدّت إلى ظهورها، كما يتعرّض لأنواع هذه الاختبارات وأغراضها في مجال التشخيص النفسي والتربوي، وفي التوجيه والإرشاد، إضافة للأغراض الإدارية وأغراض البحث العلمي ليمهّد بذلك السبيل للتعريف بعملية تصميمها وبنائها ويقف الفصل الثاني عند مسألة تصميم الاختبار النفسي وتقنينه فيشرح الأسس المعتمدة في تصميم الاختبار، وخطوات عملية تصميم الاختبار بدءاً بتحديد الهدف العام للاختبار والفكرة الموجهة له مروراً بإعداد البنود الاختبارية وانتهاءاً بالتحليل الكمي

والكيفي للبنود. ويختص الفصل الثالث بمسألة بناء الاستبانة وتوفير مستلزمات صلاحهابوصفها أداة بحث واسعة الانتشار في المجالات التربوية والنفسية، إضافة لكونهاأداة قياس وتقويم بالغة الأهمية. ولابدّ من العمل بالتالي على "استثمار" هذه الأداة على النحو الأمثل والتخلص من الممارسات الراهنة التي تظهر إساءة استعمالها في كثير من الحالات بصورة ملفتة للنظر، إن لم تكن فاضحة تماماً.

وفيما يتصل بالفصل الرابع من هذا الكتاب فقد اختص هذا الفصل بمعالجة مسائل بناء اختبارات التحصيل. سواء منها تلك التي يعدّها المعلم بنفسه لأغراض الاستعمال الصفي، أم تلك التي يعدّها الخبراء و المختصون في مجال بناء الاختبارات عادة، (بالتعاون مع خبراء ومختصين في المناهج الدراسية). ويعرف النوع الأول من هذه الاختبارات باسم "اختبارات المعلم" أو "الاختبارات الصفية" أو "اختبارات المعلم الصفية" في حين أنّ النوع الثاني منها يعرف ب"الاختبارات التحصيلية المقننة". وقد تركز الاهتمام في هذا الفصل بالتعريف بأسس بناء الاختبار الصفي، وبالتخطيط لعملية بنائه بدءاً بتحديد الغرض من الاختبار، ومن ثم تحديد الأهداف التعليمية والمحتوى الدراسي الذي سيغطيه الاختبار في نطاق ما يعرف بـ "جدول مواصفات الاختبار" مروراً باختيار شكل البنود وتقدير عددها ومستوى صعوبتها، وانتهاءاً بإعداد البنود (أو الأسئلة)، وأخيراً تحليلها، والفوائد التي يمكن أن يقدمها هذا التحليل ولاسيّما في مجال إعداد مصرف (أو بنك) شخصي للأسئلة. وقد اختتم هذا الفصل بالوقوف عند اختبارات التحصيل المقننة (أو المعيرة) والتعريف بمزاياها الخاصة بالمقارنة مع الاختبارات الصفية، وكذلك بخطوات بنائها واستخراج المعايير الخاصة بها.

أما الفصلان الخامس والسادس من هذا الكتاب فقد تناولا مسألة الصدق والثبات بوصفها المسألة المركزية في نظرية القياس المعاصرة. وقد تم التشديد في هذين الفصلين على المعنى الخاص لكل من الصدق والثبات، وعلى الأشكال أو الطرائق التي يمكن أن يأخذها كل منهما، مع الدور المسند لكل منهما في بناء اختبار يتمتع بالمواصفات الفنية المطلوبة أو الخصائص السيكومترية اللازمة التي

تظهر فاعليته كأداة قياس. هذا مع الوقوف عند معامل الثبات والخطأ المعياري للقياس إضافة إلى العوامل المؤثرة في الصدق والثبات.

والفصل السابع من هذا الكتاب يتصدّى لمعايير الاختبارات النفسية فيتعرض لمعناها وأهميتها وأنواعها المختلفة (بما فيها المعايير المئينية ومعايير الدرجات المعيارية "التائية")، كما يتعرض لدورها في رسم الصفحة النفسية للمفحوص (أو البروفيل) التي تعطي لوحة شاملة عن أدائه الاختباري، وتتيح تسليط أضواء ساطعة وقوية على نقاط ضعفه وقوته مع الوقوف أخيراً عند التحفظات التي لابدّ من مراعاتهاعند استخدام المعايير.

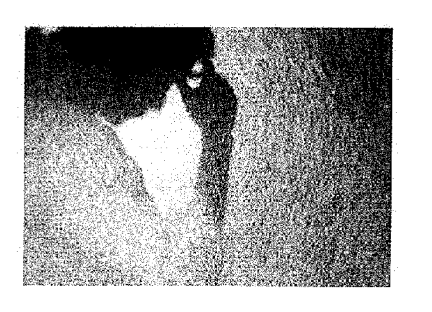
وفيما يتصل بالفصل الثامن والأخير من هذا الكتاب فقد تعرض لاستخدام الإحصاء في تحليل نتائج الاختبارات وتفسيرها انطلاقاً من أن العمل الإحصائي هو الخطوة "المواكبة" والمتمّمة لعملية بناء الاختبار، ولابدّ لمن يتولى مهمته في بناء الاختبار، أياً كان، من أن يكون قادراً على تحليل نتائجه وتفسيرها. وقد تعرّض هذا الفصل لبعض المبادئ والمفاهيم الإحصائية التي تهم العامل في القياس تحديداً، وبما يؤدّي إلى اكتساب المعارف والمهارات الإحصائية الضرورية له، ويتكامل مع المعارف والمهارت الأساسية التي تتطلبها عملية بناء الاحتبار وفق الأسس العلمية التي لا بدّ أن تستند إليها.

ولا بدّ من الإشارة إلى أنه من بين الأمور التي روعيت عند إعداد هذا الكتاب الابتعاد ما أمكن عن " الحشو" والصيغ التعبيرية المعقدة، والميل إلى التبسيط والاختصار إلى حد ما في عرض المادة العلمية على ألا يؤدي ذلك إلى السطحية والاختزال "المبتسر" وتشويه الأفكار والمعلومات. هذا مع الاهتمام بصورة خاصة بأن يأخذ الكتاب منحاه العملي ويسهم بالفعل في إكساب قارئه للمعارف والمهارات والخبرات اللازمة له لممارسة عمله في مجال تطوير وبناء الاحتبارات بدرجة عالية من الفاعلية بعد التحقق من مواصفاتها الفنية وتوافر المؤشرات السيكومترية اللازمة لها.

ويؤمل أن يقدّم هذا الكتاب بعض الخدمة والفائدة لكل من يهتم بالاختبارات النفسية، وأن يسهم بشكل أو بآخر في تطوير ممارسات أولئك الذين يسعون إلى "استثمارها" والإفادة منها في واقعنا من خلال التوسّع في استخدامها لتلبية أغراض متعدّدة، وتوظيفها على النحو الأمثل.

الفصل الأول

التعريف بالاختبارات النفسية وأنواعها وأغراضها



الفصل الأول
التعريف بالاختبارات النفسية وأنواعها وأغراضها

تحتلّ الاختبارات (أو الروائز) النفسية بأنواعها مكانة خاصة في علم النفس المعاصر، كما أن لهذه الاختبارات استعمالاتها الواسعة في مجالات حياتية عديدة ومتنوعة. ومع أن الباحثين قد يختلفون في تحديد التاريخ الدقيق لظهور حركة الاختبارات النفسية أو حركة القياس النفسي، ويتعذّر، بالتالي، تحديد نقطة زمنية معينة لبداية هذه الحركة، فإن ثمة اتفاقاً بينهم على أنها وليدة القرن التاسع عشر، وأن ثمة عوامل عديدة أسهمت في نشوئها وتطوّرها. كما أن ثمة اتفاقاً بين الباحثين على أن القياس النفسي بوصفه نهجاً جديداً يرتكز على دراسة الظاهرة النفسية من منظور علمي تجريبي بعيداً عن المنهج الاستبطاني التقليدي يمثّل تحولاً نوعياً هائلاً في تاريخ تطوّر علم النفس بل يعدّ الأساس الأهم في جعله علماً بالمعنى الدقيق للكلمة. وقد شهدت حركة الاختبارات النفسية ولاسيما في النصف الثاني من هذا القرن تطوّرات مهمة تمثّلت في ظهور بعض النظريات والأساليب الفنية، كما اتضحت أغراضها وفوائدها في المجالات المختلفة. وما من شك في أن إلقاء الضوء على تاريخ تطوّر هذه الاختبارات وأنواعها وأغراضها سيفيد في تقديم صورة – ولو أولية – عنها وعن مكانتها في الحياة المعاصرة، وسيكون بمثابة خطوة أولى لدراسة نظرية القياس المعاصرة وتطبيقاتها وفوائدها وتعرّف المكانة التي تحتلها على الصعيدين العلمي والعملي.

الخلفية التاريخية للاختبارات النفسية:

نشأة القياس النفسي وتطوّره قبل بينيه:

من العسير حصر العوامل والشروط الممهّدة لظهور حركة القياس النفسي بصورتها الحالية. فقد اختلف الباحثون في تحديد تلك العوامل وفي الوزن النسبي لكل منها بقدر اختلافهم في تعيين نقطة زمنية محددة لنشوء هذه الحركة كما ذكرنا. بيد أن الأمر الذي يكاد يجمع عليه الباحثون هو أن الدراسات الخاصة بكشف الفروق الفردية في زمن الاستجابة أو ما يسمى بدراسات "المعادلة الشخصية" التي جرت خارج إطار علم النفس وقادها الفلكيون منذ نهاية القرن الثامن عشر كانت مما مهّد في ظهور تباشير حركة الاختبارات النفسية أو ما عرف لاحقاً بحركة القياس النفسي. ففي عام 1796 حدث أن طرد أحد الفلكيين مساعده في المرصد لأنه أخطأ بمقدار ثانية واحدة في رصد الزمن الذي يستغرقه كوكب معين في مروره على لوحة التلسكوب (المرصاد الفلكي). وقد أدى هذا الحادث إلى إثارة الاهتمام بدراسة الفروق في تقديرات الفلكيين، وانتهت هذه الدراسة إلى أن الزمن اللازم لحدوث الاستجابة نحو مثير معين (زمن الرجع) يختلف من فرد لآخر، وهذا ما عرف بالمعادلة الشخصية.

من جهة أخرى بدأ الاهتمام خلال النصف الأول من القرن التاسع عشر بمسألة التخلف العقلي وبرزت الحاجة إلى التمييز بين الأفراد في القدرة العقلية. وقد قدّم الطبيب جان اسكيرول إسهاماً كبيراً في هذا المجال وذلك عندما ألحّ على ضرورة التمييز بين من أصيبوا باضطرابات عصبية وضعاف العقول مما مكّنه ولأول مرة من التمييز والفصل بين مفهومي المرض العقلي والتخلف العقلي. ولم يتوقف نشاط اسكيرول عند هذا الحد بل عمل على تصنيف مستويات التخلف العقلي، وميّز درجتين من البله Imbecility وثلاث درجات في العته Idiocy، كما لاحظ أن القدرة اللفظية واستخدام اللغة هي محك مباشر للقدرة العقلية العامة.

وقد سبق اسكيرول عصره بهذه الفكرة فاختبارات الذكاء المنتشرة حالياً مشبعة بالعامل اللفظي كما أن المحكّات المعتمدة لتشخيص حالات الضعف العقلي هي محكات لغوية في معظمها.

وتابع اسكيرول في عمله هذا سيجان الذي أنشأ أول معهد في فرنسا لتدريب ضعاف العقول عام 1837 وتركزت جهوده على إيجاد محك للتمييز بين مستويات التخلف العقلي. ويعدّ سيجان بحق أحد الرواد الأوائل الذين كرسوا اهتمامهم بضعاف العقول. وقد أورثنا سيجان لوحة الأشكال الشهيرة والمعروفة باسم "لوحة سيجان" التي تدخل ضمن الكثير من الاختبارات الأدائية وتستعمل على نطاق واسع إلى يومنا هذا في اختبار القدرة العقلية للصم والأميين ولاسيما ضعاف العقول. ولوحة سيجان ما هي إلا لوحة خشبية حفرت منها أجزاء معينة تتيح إدخال أشكال هندسية متنوعة كالمربع والمستطيل والدائرة..... إلخ. وكل ما يطلب إلى المفحوص هو وضع تلك الأشكال في الأماكن المخصصة لها.

وعموماً يمكن القول: إن أهمية الأعمال التي قدمها كل من اسكيرول وسيجان تعود إلى طرح مشكلة التخلف العقلي وتحديد مستوياته مما أدى إلى إثارة مسألة الفروق الفردية ودعا إلى الاهتمام بقياس تلك الفروق.

غير أن الدراسات الخاصة بظاهرة التخلف العقلي والدراسات التي اتجهت إلى كشف الفروق الفردية في زمن الاستجابة التي بدأها الفلكيون اقتصر دورها على خلق المناخ الملائم لظهور حركة القياس وإثارة مسألة الفروق الفردية دون أن تتعدّى هذا الدور. والعامل الحاسم في ظهور حركة القياس النفسي بوصفها حركة علمية تجريبية إلى حيز الوجود يتبدى في الاتجاه الذي قاده فونت والذي كان بمثابة نقطة تحول حاسمة في تاريخ تطوّر علم النفس وأدى إلى تجاوز المنهج الاستبطاني التقليدي ووضع حجر الأساس للمنهج التجريبي الكمي.

وقد انشأ فونت أول معمل (أو مختبر) لعلم النفس التجريبي في مدينة لايبزغ بألمانيا عام 1879. ومع أن فونت وتلامذته تجاهلوا مسألة الفروق الفردية وكانوا يهدفون من وراء تجاربهم "السيكوفيزيقية" التي تركزت على دراسة الإحساسات المختلفة من سمعية وبصرية ولمسية إلى الوصول إلى أوصاف أو قوانين عامة للسلوك البشري، فقد أسهمت محاولاتهم إسهاماً كبيراً في تشجيع حركة القياس. ذلك أن تلك التجارب تطلبت استخدام مجموعة من الاختبارات التي ركزت على الظواهر الحسية البسيطة، وكانت تلك الاختبارات بمثابة اللبنة الأولى لحركة القياس أو المؤشر الأول لظهورها. كما أن شروط الضبط التجريبي وما أملته من ضرورة توحيد الظروف المحيطة بالمفحوصين كافة، وهي ما تندرج ضمن إجراءات التقنين، انعكست بصورة مباشرة على حركة القياس ومهّدت السبيل لظهور الاختبارات المقننة التي تعدّ الإنجاز الأكبر من إنجازات هذه الحركة.

أولى فرانسيس جالتون وهو عالم البيولوجيا الإنكليزي الشهير (1822 – 1911) اهتماماً كبيراً بمسألة الوراثة عند الإنسان وأجرى دراسات موسّعة حول الصفات المختلفة عند التوائم والأقارب والأشخاص الذين لا تربطهم صلة القرابة، كما اشتهر بدراساته حول الموهوبين. ويعدّ جالتون المكتشف الحقيقي لمجال الفروق الفردية. ويمكن مقارنة دوره في حركة القياس بما انجزه وقدمه جاليله لعلم الفيزياء حيث تم بفضله التخلي نهائياً عن طرائق التقدير الحدسية والتخمينية التي استمرت لآلاف السنين والانتقال إلى طرائق علمية ترتكز على التجريب والاختبار (بوداليف، 1982).

وقد قدم جالتون مجموعة كبيرة من البيانات حول الفروق الفردية في العمليات النفسية المختلفة وصمّم بنفسه مجموعة من الاختبارات التي لازالت تستخدم حتى الآن بصورتها الأولى أو المعدلة من مثل " قضيب جالتون" للتمييز البصري للأطوال و"صفارة جالتون" لتحديد أعلى مقام سمعي وغيرها. ونظر جالتون إلى الذكاء على أنه قدرة فطرية وليست مكتسبة بالتدريب والمران وألحّ

على إمكان قياس هذه القدرة عن طريق بعض الأعمال البسيطة كما ألحّ على أن الذكاء يرتبط بالقدرة على التمييز الحسي بين الأوزان المتقاربة جداً في الوزن وذلك عن طريق اليد بدلاً من الميزان. وقد كان جالتون أول من استخدم الإحصاء في تحليل نتائج الاختبارات وأعدّ مجموعة من الطرائق والأساليب الإحصائية التي استخدمها وطوّرها فيما بعد كارل بيرسون في الولايات المتحدة الأمريكية. ويرى مورفي (Murphy,1967,p.122 نقلاً عن فرج، 1980) أن التأثير الأكبر لجالتون في حركة القياس إنما يظهر في أعمال كارل بيرسون اللاحقة التي كانت استمراراً لأعمال جالتون وطرائقه الإحصائية وتطويراً لها والأساس الذي يقوم عليه المنهج الإحصائي برمته.

وفي الولايات المتحدة الأمريكية أسهمت أعمال وجهود عالم النفس الأمريكي جيمس كاتل إسهاماً كبيراً في تطوير حركة القياس النفسي وعلم النفس التجريبي. وقد تتلمذ كاتل على يد فونت في ألمانيا ولكنه قام ببحوث متميّزة عن بحوث بقية تلاميذ فونت مهتماً بالفروق الفردية وزمن الرجع، وأعدّ رسالة عن الفروق الفردية في زمن الرجع رغم معارضة أستاذه. وقد أنشأ كاتل معملاً لعلم النفس التجريبي في امريكا وأعدّ مجموعة من الاختبارات التي تصدّت لبعض السمات الحسية الحركية من مثل قوة السمع، وحدة الإبصار، والتمييز بين الأوزان، وسرعة الحركة، والقوة العضلية، وزمن الرجع. ومع أن كاتل أعدّ هذه الاختبارات عام 1885 فإنه لم يتمكّن من نشرها إلا في عام 1896 بسبب المعارضة الشديدة التي لقيها من أستاذه فونت (بودالييف، 1982، ص28). وكان كاتل أول من استخدم مصطلح اختبار عقلي عام 1890 وذلك في معرض حديثه عن اختبارات التمييز الحسي المشار إليها.

ويعدّ كاتل بحق مؤسس حركة التجريب والقياس النفسي في الولايات المتحدة الأمريكية وزعيمها الأول. وقد أخذ عنه تلميذه العملاق كارل بيرسون الشيء الكثير. غير أن كاتل اعتقد كزميله الإنكليزي جالتون أن اختبارات التمييز الحسي وزمن الرجع تصلح لقياس العمليات العقلية الأكثر تعقيداً حيث

يمكن عن طريق بعض الأعمال البسيطة الكشف عن الفروق الفردية في القدرة العقلية بصورة غير مباشرة. ولكن سرعان ما تعرض هذا الاعتقاد الخاطئ للنقد الشديد وتزعزع بفضل الحركة الجديدة التي قادها ألفرد بينيه في فرنسا التي كانت بمثابة نقطة تحول مهمة ونقلة نوعية فريدة في تاريخ تطوّر حركة القياس النفسي إن لم تكن البداية الحقيقية لولادة هذه الحركة بصورتها الحالية ونموها وازدهارها إلى يومنا هذا.

بينيه وقياس الذكاء:

يعدّ ألفرد بينيه المؤسس الحقيقي لحركة قياس الذكاء بصورتها الحالية. والواقع أن بينيه انشغل منذ أواخر القرن التاسع عشر بمحاولة إيجاد أداة مناسبة لقياس الذكاء وقد أعدّ بالاشتراك مع هنري مقالاً نشر في عام 1895 وجّه فيه نقداً شديداً إلى الاتجاه السائد آنذاك في القياس العقلي. وتركزت أفكار بينيه على أن الفروق في إدراك المثيرات الحسية لا تعبر بالضرورة عن الفروق الحقيقية في الذكاء، وأن اختبارات التمييز الحسي وزمن الرجع لا تصلح لقياس العمليات والوظائف العقلية العليا كالتذكر والانتباه والتفكير، ولا مناص بالتالي من قياس هذه الوظائف والعمليات العقلية المعقدة بصورة مباشرة ودون الاعتماد على المقاييس الحسية والحركية السابقة.

وفي عام 1904 أصدر وزير التعليم الفرنسي قراراً بتشكيل لجنة أوكلت إليها مهمة إيجاد الوسيلة أو الأداة المناسبة لعزل الأطفال المتخلفين عقلياً من البرامج التعليمية المدرسية. واقترح بينيه وزميله سيمون عضوا هذه اللجنة مقياسهما الأول للذكاء الذي عرف باسم مقياس بينيه – سيمون، فظهر إلى الوجود أول مقياس للذكاء بالمعنى المعروف حالياً عام 1905.

ضم المقياس بصورته الأولى ثلاثين بنداً مرتبة تصاعدياً بحسب درجة صعوبتها، وتمّ تحديد مستوى الصعوبة بتطبيق البنود على 50 طفلاً سوياً تتراوح

أعمارهم بين 3 سنوات و11 سنة، وعلى بعض الأطفال المتخلفين عقلياً. وأجرى بينيه تعديلاً على هذا المقياس عام 1908 فزاد عدد البنود وحذف ما لم تثبت صلاحيته منها، وجمع هذه البنود وصنّفها في مستويات عمرية فوضعت في مستوى 3 سنوات البنود التي يستطيع الطفل العادي (الطفل المتوسط) في سن الثالثة الإجابة عنها بنجاح، وهكذا في بقية الأعمار حتى الثالثة عشرة. وطرح بينيه مفهوم العمر العقلي والذي يعبر عن أداء الطفل المتوسط في عمر زمني معين، وأمكن عن طريق هذا المعيار مقارنة درجات أي طفل في الاختبار بمتوسط درجات أبناء عمره، كما أمكن مقارنة أداء الطفل بالمستويات العمرية الأدنى والأعلى من عمره، فإذا استطاع طفل في الخامسة من عمره اجتياز الاختبارات الخاصة بالمستوى العمري 6 سنوات فإن عمره العقلي هو 6 سنوات، وإذا لم يتمكن طفل في الثامنة من اجتياز الاختبارات المخصصة للمستوى العمري 8 سنوات وتمكن من اجتياز الاختبارات المخصصة للمستوى العمري 7 سنوات فإن عمره العقلي هو 7 سنوات. وكان لمعيار العمر العقلي أهمية فائقة بوصفه محكاً للتمييز بين مستويات الذكاء، وجذب الاهتمام لاختبارات الذكاء لسهولته ووضوح دلالته.

وفي عام 1911 اجرى بينيه تعديلاً آخر على مقياسه وأضاف إليه بنوداً جديدة لمستويات عمرية أعلى تصل إلى عمر الراشدين. وقد لاحظ شترن، وأيّده في ذلك تيرمان، تغيّر العمر العقلي للطفل مع تقدّمه في العمر الزمني، وأشار إلى أن مفهوم العمر العقلي يعبّر بصورة مطلقة عن تقدّم أو تخلف الفرد مقيساً بالسنوات دون نسبة هذا التقدّم أو التخلف إلى عمره الزمني. ونتيجة لذلك تمّ إدخال مفهوم جديد إلى القياس العقلي وهو مفهوم حاصل الذكاء (نسبة الذكاء) والذي يحسب بنسبة العمر العقلي إلى العمر الزمني ثم يضرب بـ (100) ويعبر عن التقدّم أو التخلف النسبي للفرد من خلال مقارنة أدائه الفعلي، كما يعبر عنه عمره العقلي، بعمره الزمني. وبخلاف العمر العقلي الذي قد يتغير بصورة ملحوظة مع التقدّم في العمر الزمني فإن حاصل الذكاء (نسبة الذكاء)

يبقى ثابتاً، أو يتغير قليلاً، من سنة لأخرى بالنسبة لأغلب الأطفال (Murphy,1927,p.535 نقلاً عن: فرج، 1980).

أثار مقياس بينيه – سيمون الفرنسي للذكاء اهتماماً كبيراً لدى الأوساط المعنية في الولايات المتحدة الأمريكية. وفي عام 1910 بدأ لويس ترمان بإعداد دراسة موسّعة حول هذا المقياس في جامعة ستانفورد وطبّقه على عينات أمريكية تتألف من 2300 طفلاً. وتضمّنت هذه الدراسة تعديلات كبيرة على المقياس، وانتهت إلى نشر المقياس بصورته المعدلة عام 1916 باسم مقياس ستانفورد – بينيه، وأصبح المقياس بذلك مقياساً أمريكياً من أصل فرنسي إن جاز التعبير. وسرعان ما انتشر هذا المقياس على نطاق واسع في الولايات المتحدة وغيرها من البلدان. وقد استطاع هذا المقياس أن يثبت وجوده على مرّ الزمن بوصفه مقياساً للذكاء على درجة عالية من الجودة. وأسهمت التعديلات العديدة التي تعرض لها فيما بعد في تحسينه وتعزيز مكانته وما زال يحتفظ بقيمته ويتبوأ مكانة خاصة حتى يومنا هذا.

تطوّر قياس الذكاء بعد الحرب العالمية الأولى:

كان للحرب العالمية الأولى تأثيرها الحاسم في دفع حركة قياس الذكاء، فقد برزت الحاجة في أثناء تلك الحرب إلى استخدام مقاييس عقلية جمعية تطبق على أعداد كبيرة من الأفراد في وقت واحد، واستدعى ذلك تصميم اختبارات ذكاء جمعية سهلة التطبيق كان أشهرها اختبار ألفا للجيش واختبار بيتا للأميين وغير الناطقين بالإنكليزية. وقد استخدمت نتائج هذه الاختبارات في فرز المجندين وانتقاء القادة ورجال المهمات الخاصة كما استخدمت في استبعاد المتخلفين عقلياً وتحويل منخفضي الذكاء إلى أعمال مناسبة. وهذا الحماس الذي شهدته حركة القياس العقلي في أثناء الحرب العالمية الأولى لم يخمد في فترة السلم. وبدءاً من ثلاثينات القرن الماضي انتشرت مقاييس الذكاء في بلدان عديدة واستخدمت لأغراض عديدة. ففي فرنسا استخدمت مقاييس الذكاء

لأغراض التوجيه المهني والتشخيص النفسي. وفي الولايات المتحدة وإنكلترا دخلت الاختبارات إلى مختلف مجالات الحياة، واعتمدت نتائجها في دخول المدارس والمعاهد والجامعات وفي الحصول على عمل أو وظيفة.... إلخ. وفي الاتحاد السوفيتي استخدمت اختبارات الذكاء في مجالي التعليم والتوجيه والاصطفاء المهني.

أدى هذا الانتشار الواسع للمقاييس العقلية وبفعل تأثيرها المباشر في حياة الملايين من الناس إلى ظهور تيارين متعارضين: الأول يناصر حركة القياس والثاني يعارضها. ففي عام 1936 أصدرت الجهات المسؤولة في الاتحاد السوفيتي قراراً بحظر استخدام الاختبارات العقلية من منطلق أنها ليست أدوات علمية صادقة، كما أنها يمكن أن تؤدي إلى تعزيز الفروق الطبقية وتكريس التمايز الاجتماعي بين الناس. وهذا ما أدى إلى جمود حركة القياس وتعطيلها تماماً نحو ثلاثين عاماً في الاتحاد السوفيتي (ليونتيف، 1968)، كما وجّهت إلى الاختبارات العقلية انتقادات شديدة في أرجاء أخرى من العالم بما في ذلك الولايات المتحدة الأمريكية (المصدر نفسه).

غير أن الهجوم الشديد على اختبارات الذكاء لم يوقف مسار هذه الحركة ونموها. ففي عام 1937 حدث تطوّر مهم في حركة القياس حيث ظهر التعديل الجديد لمقياس ستانفورد – بينيه من قبل تيرمان وميريل، وقد تمّ تقنين هذا المقياس على عينات واسعة من الأفراد ورفع فيه سقف العمر العقلي إلى 22 سنة، وأصبح له صورتان هما الصورة L والصورة M. وحظي هذا المقياس الجديد بانتشار واسع واستمر العمل به حتى أوائل السبعينيات حيث ظهر تعديل دمَجَ الصورتين في صورة واحدة.

وفي عام 1939 حدث تطوّر مهم آخر في حركة القياس. فقد وضع ديفيد فكسلر المختص النفسي في مستشفى بلفيو في نيويورك مقياساً للذكاء عرف باسم مقياس فكسلر – بلفيو. وجاء هذا المقياس ليلبي الحاجة إلى مقياس ذكاء

للراشدين، كما استخدم على نطاق واسع للأغراض التشخيصية. ويتضمّن مقياس فكسلر – بلفيو مجموعتين من الاختبارات الفرعية، الأولى هي مجموعة الاختبارات اللفظية، والثانية هي مجموعة الاختبارات الأدائية. ويفيد تطبيق هاتين المجموعتين في استخلاص حاصل الذكاء اللفظي وحاصل الذكاء الأدائي وحاصل الذكاء الكلي (العام). وتبرز أهمية هذا المقياس في أنه يولي اهتماماً كبيراً للجوانب الأدائية (غير اللفظية) للذكاء، ويسدّ بذلك فجوة كبيرة تعاني منها مقاييس الذكاء المشبعة بالعامل اللفظي في معظمها، كما يتميّز بخصائصه التمييزية الدقيقة وقدرته التشخيصية العالية. وقد وضع فكسلر بالإضافة لهذا المقياس مقياساً خاصاً بالأطفال من 5 – 15 سنة، كما طوّر مقياساً للأطفال من 4 – 6 سنوات ونصف، وسار فكسلر في تصميمه لهذين المقياسين على النهج نفسه الذي اتبعه في تصميم مقياسه الأول.

وكان لمنهج التحليل العاملي دور لا يستهان به في تطوّر حركة القياس العقلي فقد استخلص سبيرمان منذ أوائل القرن الماضي عامل الذكاء العام (العامل العام) بوساطة التحليل العاملي لنتائج مجموعة من الاختبارات. وجاء رفض ثرستون لنظرية سبيرمان في العامل العام الذي يتضمّن أن نتائج الارتباطات بين مجموعة من الاختبارات لا تعبر عن عامل واحد عام بل تعبر عن عوامل متعدّدة، بمثابة قوة دافعة جديدة لحركة القياس. ووفقاً لنظرية العوامل المتعدّدة صمّم ثرستون بطارية القدرات العقلية الأولية PMA. وانتشرت هذه البطارية على نطاق واسع خلال الثلاثينيات، وتضمّ هذه البطارية اختبارات لسبع قدرات أولية هي: الفهم اللفظي، والطلاقة اللفظية، والقدرة العددية، والإدراك المكاني، والتذكر، والسرعة الإدراكية، والاستدلال المنطقي. ويعطي تطبيق هذه البطارية من الاختبارات درجات منفصلة لكل قدرة على حدة مما يفسح المجال لتحليل قدرات الفرد وتشخيصها، كما "يقدّم تصويراً أكثر تركيباً وتعقيداً لقدرات الفرد من الدرجة الكلية التي يقدّمها بينيه أو التصنيف الثنائي الذي يتضمّنه مقياس فكسلر" (فرج، 1980، ص32). وقد عارض تومسون فكرة العامل العام

كما عارض فكرة العوامل المتعدّدة. غير أن بحوثه في مجال التحليل العاملي كان لها وزنها، وأدت هذه البحوث، كما أدت بحوث زملائه من قبله إلى تطوير البحوث الخاصة بالقدرات العقلية وقياسها. وعموماً فإن من الملامح البارزة لحركة القياس أثناء الحرب العالمية الثانية وبعدها الاهتمام باختبارات القدرات الخاصة التي بدأت تنافس اختبارات الذكاء العام وكانت بمثابة خطوة متقدّمة في حركة القياس. وقد جاءت هذه الاختبارات لتلبي حاجات عملية ملحة برزت أثناء تلك الحرب الساخنة واشتدت في فترة الحرب الباردة التي أعقبتها. ويمكن توقع زيادة الاهتمام باختبارات القدرات الخاصة في الفترة التي يشهدها عالمنا الآن والتي تتّسم بتنافس سلمي شديد، دون أن يؤدي ذلك بالضرورة إلى التقليل من أهمية اختبارات الذكاء العام والانتقاص من شأنها أو حذفها من الوجود.

قياس الشخصية:

من الملامح البارزة لحركة القياس ظهور الحركة النشطة في مجال قياس الشخصية التي بدأت في أثناء الحرب العالمية الأولى وانتشرت بعدها على نطاق واسع. والواقع أن حركة قياس الشخصية، لم تظهر بصورة مفاجئة في أثناء الحرب العالمية الأولى بل كانت نتاج محاولات وجهود سابقة واستمراراً لها تماماً كحركة قياس الذكاء. ومما مهّد لظهور هذه الحركة الجهود التي بذلها كارل بيرسون وروزانوف في الولايات المتحدة ويونغ في سويسرا وكريبلن في ألمانيا، وكان الهدف من وراء تلك الجهود والمحاولات هو الكشف عن سمات الشخصية والتمييز بين الأفراد الأسوياء وغير الأسوياء. من جهة أخرى أدى انتشار اختبارات الذكاء العام والقدرات الخاصة واستخدامها على نطاق واسع في مجالات الحياة المختلفة إلى إثارة العديد من الأسئلة بصدد فاعلية هذه الاختبارات وقدرتها على الكشف عن مستوى الأداء العقلي للفرد حيث تبين أن ثمة عدداً من المتغيرات "غير العقلية" التي يمكن أن تتدخل في الأداء العقلي للفرد، وتتصل هذه المتغيرات بشخصية الفرد وسماته المزاجية مما مهّد بدوره لتشجيع الحركة الخاصة بقياس الشخصية وعدم الاقتصار على الاختبارات العقلية وحدها في دراسة الفرد.

قياس الشخصية عن طريق التقرير الذاتي:

من الأسماء المهمة في حركة قياس الشخصية روبرت وودورث الذي وضع في أثناء الحرب العالمية الأولى قائمة لقياس الشخصية معتمداً أسلوب التقرير الذاتي أي وصف الشخص لنفسه وتقديره لسمات شخصيته وانفعالاته الخاصة. وتتضمّن قائمة وودورث للشخصية 116 سؤالاً وتتناول مجموعة من اضطرابات الشخصية كالمخاوف والوساوس، والكوابيس وغيرها. وقد طبقت هذه القائمة على أعداد كبيرة من المجندين بهدف التعرّف على المضطربين نفسياً والذين يعانون من مشكلات انفعالية حادة تؤثّر في أدائهم في الحرب. وفي عام 1928 ظهر اختبار ألبورت للسيطرة والخضوع والذي اعتمد بدوره أسلوب التقرير الذاتي واتجه إلى الكشف عن الميل إلى السيطرة أو الخضوع من خلال دراسة استجابات الفرد للمواقف الاجتماعية التي تتأثر بذلك الميل.

غير أن التطوّر الحاسم في حركة قياس الشخصية عن طريق التقرير الذاتي تأكد مع ظهور اختبار مينيسوتا المتعدّد الأوجه للشخصية MMPI الذي وضعه هاتاواي وماكينلي عام 1943. وتضمّن هذا الاختبار 550 بنداً تتناول الجوانب المختلفة للشخصية، كالعادات الشخصية، والمواقف من الدين، والسياسة، والمجتمع، والسمات الانفعالية وغيرها. وقنّن هذا الاختبار على عينة من الأسوياء والمرضى المقيمين في مشافي الأمراض العقلية. ويمكن من خلال تطبيق هذا الاختبار رسم الصفحة النفسية للفرد (البروفيل) التي تعبر بوضوح عن درجات في الاستواء أو الاضطراب النفسي لدى الفرد في المقاييس المختلفة التي يضمّها الاختبار. ويعاب على اختبارات الشخصية عن طريق التقرير الذاتي أن البيانات التي تقدّمها قد لا تصف الشخص بصورة موضوعية وصادقة طالما أنها تعتمد على وصف الشخص لنفسه بنفسه. وقد ظهر أن المفحوصين يميلون غالباً إلى الرد بصورة إيجابية على الأسئلة التي تتضمّن بعض الخصائص أو المواقف الاجتماعية المرغوبة، كما أن ردودهم غالباً ما تكون سلبية تجاه النواحي غير المرغوبة اجتماعياً.

الطرائق الإسقاطية:

والاتجاه الثاني في حركة قياس الشخصية تمثل في الطرائق الإسقاطية التي بدأها الطبيب النفسي السويسري هرمان رورشاخ منذ عام 1911 والذي لاحظ أن الأسوياء والمرضى من الناس عندما ينظرون إلى اشكال غامضة لا معنى لها كبقع الحبر مثلاً فإنهم يرون فيها أشياء عديدة ومختلفة مما يشير إلى أنهم يسقطون اتجاهاتهم وعقدهم الشخصية على تلك المثيرات الغامضة. وقد وضع رورشاخ اختباره الشهير لبقع الحبر عام 1921 ويحتوي هذا الاختبار على عشر بطاقات بيضاء تظهر على كل منها بقعة كبيرة من الحبر وتأخذ شكل نصفين متناظرين. وتكون مهمة المفحوص هي التعبير عما يراه في تلك البقع بحريّة، وتؤخذ استجاباته دليلاً على بعض سماته الشخصية الانفعالية. ويعدّ هذا الاختبار من أكثر الاختبارات الإسقاطية شيوعاً وانتشاراً ويستخدم بخاصة لأغراض التشخيص الإكلينيكي.

ومن الاختبارات المهمة التي اعتمدت الطرائق الإسقاطية اختبار تفهم الموضوع الذي نشره موراي ومورجان عام 1935. وتضمّ المجموعة الكاملة لهذا الاختبار ثلاثين بطاقة أو صورة تعبّر عن مواقف مختلفة كالخطر والخوف والعدوان والميول الانتحارية والجنس... إلخ، بالإضافة إلى بطاقة واحدة بيضاء. ويطلب إلى المفحوص أن يؤلف قصة من خياله تناسب كل صورة من الصور المعروضة عليه. ويقوم هذا الاختبار على افتراض أن المفحوص سيسقط مشاعره الذاتية وقيمه واتجاهاته على تلك المثيرات كما سيعبّر عن الضغوط التي يعاني منها. ويختلف هذا الاختبار عن اختبار بقع الحبر في أنه يتضمّن مواقف شبيهة بمواقف الحياة الواقعية ولذا فهو ينتمي إلى فئة الاختبارات المحددة البناء جزئياً. ويستعمل هذا الاختبار في البحث وفي التشخيص الإكلينيكي.

وعموماً توجّه إلى الاختبارات الإسقاطية انتقادات شديدة منها أن هذه الاختبارات تضمّ مثيرات غامضة وغير موحّدة المعنى، كما أن "معايير" تفسير

الدرجات فيها كثيراً ما تتأثر بالعوامل الذاتية للمختبر مما يطرح سؤالاً كبيراً حول تقنينها وصحة معاييرها، هذا بالإضافة إلى أن تلك الاختبارات تعاني من ضعف مستوى الصدق والثبات (الموثوقية) بصورة واضحة مما يدعو إلى التساؤل حول جدواها كأدوات بحث وقياس، والتشكيك بقيمتها العلمية والعملية.

مقاييس الاتجاهات والميول:

بالإضافة إلى طرائق الإسقاط والتقرير الذاتي التي عملت على دراسة الشخصية بوصفها كلاً متكاملاً، ظهر اتجاه لدراسة جوانب ومتغيرات مهمة في الشخصية كالاتجاهات (أو المواقف)، والميول، والقيم، والآراء بهدف إلقاء المزيد من الضوء على الشخصية بأبعادها المختلفة. والواقع أن الاتجاه إلى دراسة الاتجاهات والميول وغيرها جاء متمّماً للاتجاهات السابقة التي تركزت على دراسة السمات الأساسية للشخصية. واستطاع هذا الاتجاه أن يطوّر أدوات بحث وقياس على درجة عالية من الدقة متخطياً بذلك الكثير من الصعوبات المنهجية التي تعرضت لها تلك الاتجاهات.

ومن المحاولات المبكرة في هذا المجال المحاولة التي قام بها ثيرستون في بداية العقد الثالث من هذا القرن والذي وضع مع مساعديه نحو ثلاثين سلماً للاتجاهات تناولت موضوعات عديدة كالمواقف من الحرب، ومن الزنوج، ومن الشيوعية، والرأسمالية، ومن الكنيسة...إلخ. وقد انتشرت مقاييس ثيرستون على نطاق واسع واستخدمت الطريقة التي اتبعها التي اعتمدت آراء المحكّمين، في تطوير العديد من مقاييس الاتجاهات. كما وضع ليكرت عدداً من مقاييس الاتجاهات شملت العديد من القضايا، وسلك طريقة جديدة في بناء هذه المقاييس مخالفاً طريقة ثيرستون التي تعتمد على آراء المحكّمين في تقويم البنود المختلفة. واعتمدت طريقة ليكرت على تقديم جمل محايدة وخمسة بدائل للإجابة تعبر عن شدة الاتجاه. وهذه الطريقة أبسط من طريقة ثيرستون وأكثر شيوعاً وانتشاراً.

وظهرت المحاولات الأولى لدراسة الميول وقياسها منذ أوائل القرن الماضي. ففي عام 1907 وضع استانلي هول استبانة الكشف عن ميول الأطفال نحو فعاليات الاستجمام. وتبعت هذه المحاولة محاولات أخرى كان أبرزها مقياس سترونغ للميول المهنية الذي ظهر في عام 1927 وتضمّن عدداً كبيراً من البنود. واتجه هذا المقياس إلى الكشف عن الميول الخاصة بكل من الجماعات المهنية المختلفة كالمعلمين والأطباء والمهندسين وغيرهم. وقد حظي باهتمام كبير وانتشر على نطاق واسع وتمّ تعديله وتحسينه عام 1974 حيث ظهر باسم استخبار سترونغ – كامبل للميول. ولهذا المقياس أهمية خاصة في مجال التوجيه المهني واختيار المهنة. ومن الأدوات المهمة لقياس الميول قائمة كودر للميول والتفضيلات التي ظهرت بصورتها الأولى عام 1939 واستخبار لي – ثورب للميول المهنية.

وعموماً تقدّم مقاييس الاتجاهات والميول إسهاماً كبيراً في مجال دراسة تلك الجوانب والمتغيرات المهمة في الشخصية التي لا تتعرض لها مقاييس الشخصية ذات الطبيعة الكلية والشاملة. ولهذه المقاييس دورها في تطوير معارفنا حول الشخصية وإغنائها. كما أنها تقدّم خدمات عديدة في مجال الحياة العملية ولاسيما في مجال التوجيه المهني واختيار المهنة.

أنواع الاختبارات النفسية:

ما الاختبار النفسي؟

من التعريفات الشائعة في أدبيات القياس النفسي للاختبار (أو الرائز) النفسي التعريف الذي تقدّمه آنا أنستازي والذي تقول فيه:" إنه مقياس موضوعي ومقنن لعينة من السلوك " (Anastasi,1982,p.22). ويتّضح من هذا التعريف أن الاختبار النفسي يجب أن يعطي تقديراً موضوعياً لأداء المفحوص ولا يفسح المجال لظهور اختلافات بين الفاحصين أو لدى الفاحص الواحد في ظروف وأوقات

مختلفة مما "يلغي" أثر العوامل الذاتية أو الشخصية في الحكم على أداء المفحوص أو بالأحرى يخفّف من وطأتها إلى الحدود القصوى المتاحة. ومن المفيد الإشارة ههنا إلى أن الموضوعية يمكن أن تكون خاصية لنتيجة عملية القياس ولا تكون بالضرورة خاصية للعملية التي تؤدي إلى تلك النتيجة بكل ما فيها من إجراءات وخطوات. وهذا يعني أن عملية وضع الدرجات فقط (Scoring) أو ما يعرف بتصحيح الاختبار يمكن أن تتمّ بطريقة موضوعية، وأما عملية تصميم الاختبار فلابد أن تتأثر بالعوامل الذاتية. فإذا قام باحثان بوضع اختبارين "موضوعيين" يفترض أنهما يقيسان شيئاً واحداً على سبيل المثال وطبّقاهما على مجموعة واحدة من المفحوصين فيمكن أن نتوقع حصول المفحوصين على درجات مختلفة، على الرغم من التصحيح الموضوعي لهذين الاختبارين.

غير أن تعريف أنستازي يلحّ أساساً على التقنين، والتقنين في جوهره يعني التوحيد، ويتضمّن توفير شروط واحدة لجميع المفحوصين في تطبيق الاختبار ووضع درجاته مما يستدعي توحيد التعليمات والأمثلة التمهيدية والحدود الزمنية وطريقة الإجابة والشروط المحيطية كالإنارة والتهوية...إلخ، ويتيح بالتالي ضبط العوامل والمتغيرات جميعها التي يمكن أن تؤثّر في الأداء الاختباري. كما يلحّ هذا التعريف أيضاً على أن الاختبار النفسي ما هو إلا عينة من المثيرات يمكن من خلالها الحصول على عينة من الاستجابات أو عينة من سلوك المفحوص. وبطبيعة الحال فإن عينة الاستجابات المشار إليها لابد أن تكون ممثلة للمجتمع الأصلي للاستجابات (أو المجتمع الأصلي للسلوك) الذي يغطي السمة أو الخاصية المقيسة وإلا فإن الأداء الاختباري للمفحوص لن يكون دليلاً صادقاً ومعبراً بدقة عن أدائه الحقيقي فيما يتصل بقطاع السلوك الذي تظهر من خلاله السمة أو الخاصية موضع الاهتمام.

ومن التعريفات المهمة للاختبار النفسي التعريف الذي يقدّمه كرونباش والذي يقول فيه: "إنه طريقة منظمة لمقارنة سلوك شخصين أو أكثر" (Cronbach,1960,p.21). ولا يتعارض هذا التعريف مع تعريف أنستازي

السابق، وقد لا يختلف عنه في جوهره، ويصح على الأرجح، أن نقول إنه يتكامل معه في تقديم صورة صادقة عن الاختبار النفسي. فتعريف كرونباش ينطوي بصورة ضمنية على ضرورة توفر شرط التقنين في الاختبار النفسي لكي يتيح " مقارنة سلوك شخصين أو أكثر" بصورة سليمة. كما يؤكد صراحة أنه "طريقة منظمة" مما يستبعد احتمال التسرع والعشوائية والارتجال في تصميمه وإجرائه وتفسير نتائجه. غير أن هذا التعريف يتراجع قليلاً عن النظرة الصارمة إلى الاختبار النفسي التي تجعله أداة قياس كمية على درجة عالية من الموضوعية والدقة، ويفسح المجال لأدوات التقدير غير الكمية لتحتل مكانها بين الاختبارات النفسية. إذ من المعلوم أن الكثير من اختبارات الشخصية مثلاً تكتفي بأوصاف وتقديرات كيفية لأداء المفحوص ولا تعطي تقديرات كمية رقمية. كما أن هذا التعريف يؤكد دور الاختبار النفسي في المقارنة بين الأفراد على حين أن التعريف السابق يلحّ على الاختبار بوصفه عينة من المثيرات اللفظية أو غير اللفظية تؤدي بطبيعة الحال إلى ظهور عينة من الاستجابات يفترض أن تكون صادقة وممثلة لقطاع واسع من السلوك.

تصنيف الاختبارات النفسية:

نادراً ما تصنّف الاختبارات أو الروائز النفسية في معزل عن المقاييس التربوية أو أدوات التقويم التربوي المختلفة. وتميل أغلب أدبيات القياس النفسي والتقويم التربوي إلى النظر إليها مجتمعة بوصفها " أسرة" واحدة. ومن التصنيفات الشائعة لتلك الاختبارات والأدوات التصنيف الذي ينطلق من الموضوعات أو الصفات أو المظاهر السلوكية التي تتصدّى لها. فإذا أخذنا بهذا المنطلق في التصنيف واعتمدناه أساساً في تصنيف الاختبارات النفسية توصلنا إلى عدد من الأصناف او الأنواع التي يرتبط كل منها بالصفة أو الظاهرة النفسية التي نخضعها للقياس. وتبعاً لذلك يمكن التمييز بين اختبارات الذكاء العام، واختبارات القدرات الخاصة، بالإضافة إلى مقاييس الشخصية، وقوائم (أو استخبارات) الاتجاهات، والميول، والقيم، واستطلاعات (أو كشوف) الآراء.

ومن التصنيفات المهمة للاختبارات النفسية تصنيف كرونباش الذي يضعها ضمن فئتين أو صنفين رئيسين اثنين هما:

1. اختبارات أقصى الأداء Tests of Maximum Performance:

وتطلب هذه الفئة من الاختبارات إلى المفحوص تقديم أفضل أو أجود ما لديه للحصول على أعلى درجة، وتستعمل للكشف عن مستوى القدرة أو أقصى الأداء الذي يمكن أن يصل إليه. وتتضمّن هذه الفئة اختبارات القدرات والاستعدادات المختلفة. ومن أمثلتها اختبارات القدرة العقلية العامة أو اختبارات الذكاء العام التي تتصدّى لقياس "تشكيلة " واسعة من المهمات والأنشطة العقلية وتهدف إلى تحديد المستوى العقلي العام للمفحوص وتعبر عنه من خلال رقم حاصل الذكاء (أو نسبة الذكاء)، واختبارات القدرات الخاصة التي تتصدّى لقياس قدرات نوعية محددة كالقدرة اللفظية، والقدرة الميكانيكية، والقدرة العددية، وغيرها. وتتضمّن اختبارات القدرات الخاصة مجموعة متجانسة أو "طائفة" محددة من المهمات العقلية وتسمى أحياناً اختبارات القدرات الطائفية. وكثيراً ما تتداخل هذه الاختبارات مع اختبارات الاستعدادات الفارقية Differential Aptitude Tests التي تسعى إلى الكشف عن الفروق في ذات الفرد في سمات معينة كالفهم اللفظي والاستدلال الحسابي والاستدلال الميكانيكي. وتتيح اختبارات الاستعدادات الفارقية رسم الصفحة النفسية للفرد (البروفيل) مما يفيد في التشخيص الفارقي أو تحليل ذات الفرد وإجراء مقارنات بين القدرات أو الاستعدادات الخاصة للفرد الواحد. كما تنتمي إلى فئة اختبارات أقصى الأداء اختبارات التحصيل الدراسي بأنواعها، وبطبيعة الحال تلك الاختبارات التي يختلط فيها التحصيل بالاستعداد أو اختبارات التحصيل والاستعداد Combination Aptitude – Achievement Tests.

2. اختبارات الأداء النمطي أو الأداء العادي

Tests of Typical Performance:

وهي تستخدم لمعرفة ما يحتمل أن يفعله الشخص في موقف معين وكيف يتصرف في الأوضاع العادية وليس لمعرفة ما يستطيع أن يفعله. وتبعاً لذلك لا تسعى هذه الفئة من الاختبارات إلى الكشف عن أعلى مستوى من الأداء يمكن أن يصل المفحوص إليه ولا تعطي تقويماً لاستجاباته بوصفها صحيحة أو خاطئة أو جيدة أو رديئة بل تهتم برصد السمات أو الخصائص السلوكية كما هي عليه. وتدخل ضمن هذه الفئة من الاختبارات اختبارات الشخصية الكلية واختبارات الاتجاهات والميول والآراء والقيم. ويمكن تصنيف اختبارات الأداء النمطي إلى فئتين هما:

1. أدوات الملاحظة: وتقوم على ملاحظة الآخر لسلوك الفرد، وتشمل هذه الأدوات الاختبارات الموقفية التي تقوم على الملاحظة المباشرة لسلوك الفرد في مواقف مدبرة بصورة مسبقة. كما تشمل قائمة الرصد (أو قائمة الشطب) وتستعمل في تقويم الميول والاتجاهات بالإضافة إلى الصفات الشخصية وبعض مظاهر التكيف الشخصي والاجتماعي، كما تشمل أيضاً سلم الرتب الذي يتميّز عن قائمة الرصد في احتوائه على درجات للصفة أو الخاصية المعنية ويستعمل بدوره في تقويم سمات الشخصية ومظاهر التكيف الشخصي والاجتماعي.
2. تقنيات التقرير الذاتي: وتستهدف الحصول على معلومات معينة عن الفرد من الفرد ذاته بصورة مباشرة. ومن أدوات التقرير الذاتي الاستبانة أو الاستخبار Questionnaire وهي قائمة من الأسئلة التي تتناول بعض الصفات أو الموضوعات التي تتعلق بحياة الفرد ومشاعره ومعتقداته وتفضيلاته وميوله المهنية واتجاهاته. وكثيراً ما يطلق على هذا النوع من الأدوات اسم القوائم Inventories ولاسيما عندما تستعمل لقياس

الشخصية والميول، ويستعمل تبعاً لذلك مصطلح قوائم الشخصية أو قوائم الميول للإشارة إليها. وعيب القوائم والاستبانات هو احتمال تزييف الإجابة بهدف الظهور بالمظهر الملائم مما أدى إلى استعمال مقاييس خاصة للكشف عن الكذب والتزوير كما هو الحال في مقياس مينيسوتا متعدّد الأوجه للشخصية.

وثمة تصنيفات أخرى عديدة للاختبارات النفسية لا يتسع المقام للإحاطة بها. ونكتفي بالوقوف عند التصنيفات التالية لشيوعها وأهميتها:

تصنيف الاختبارات انطلاقاً من شروط إجرائها:

المقصود بشروط الإجراء تلك الشروط الخاصة بتطبيق الاختبار على المفحوصين إما فرادى أو جماعات. وفي هذا الإطار يمكن التمييز بين الاختبارات الفردية والاختبارات الجمعية. ومن الفروق المهمة بين الاختبارات الفردية والاختبارات الجمعية أن التعليمات في هذه الأخيرة تكون أبسط منها في الأولى، كما لا تتطلب الاختبارات الجمعية خبرة أو مهارة خاصة من جانب الفاحص، وهي أقل كلفة وأكثر انتشاراً من الاختبارات الفردية. غير أن الاختبارات الفردية تعطي الفرصة للتواصل الحي بين الفاحص والمفحوص مما يفسح المجال لاستثارة دافعية المفحوص وضمان ملاحظة الطرائق التي يلجأ إليها المفحوص في أدائه الاختباري وأساليب حله للمشكلات والتعرّف على نوع الأخطاء التي يرتكبها.

ومن شروط الإجراء ما يتصل بطريقة المفحوص في التعبير عن نفسه وإنجازه في الموقف الاختباري كأن يعبر عن نفسه شفهياً أو كتابة، أو بالكلمات والألفاظ، أو بالأداء. وهنا يمكن التمييز بين الاختبار اللفظي واختبار الأداء، كما يمكن التمييز بين الاختبار الشفهي والاختبار الكتابي.

ويتميّز الاختبار اللفظي باعتماده على فهم واستخدام الكلمات، ويشغل المرتبة الأولى بين أدوات القياس من حيث أهميته واستخدامه في المجالات كافة

مما يشير إلى المكانة التي يحتلها الأداء اللفظي والأهمية الفائقة للقدرة اللفظية في ثقافتنا وحضارتنا المعاصرة. وقد يعتمد الاختبار اللفظي اللغة المكتوبة في الإجابة ويطلق عليه في هذه الحالة اختبار الورقة والقلم. وقد يعتمد اللغة المنطوقة أو التعبير الشفهي من جانب المفحوص (الاختبار الشفهي). وهو في الحالتين يتميّز عن اختبار الأداء الذي يعتمد غالباً على الأداء العملي ومعالجة الأشياء يدوياً كالصور أو المكعبات أو الأدوات. ولابد من التمييز في نطاق هذا النوع الأخير من الاختبارات بين اختبارات التحصيل العملية أو الأدائية التي تقيس أنواعاً معينة من الأداء المتعلّم كالأداء الموسيقي والأداء الرياضي وتحضير الطعام والأعمال المخبرية وتشغيل الآلات والأدوات وغيرها، واختبارات الاستعدادات أو القدرات العملية التي كثيراً ما تسمى اختبارات الذكاء العملي أو القدرة العملية. وتجدر الإشارة إلى أن الاختبارات الأدائية من النوع الأخير قد تنتمي إلى فئة الاختبارات غير اللغوية Nonlanguage Tests التي تستخدم مع الأجانب أو العجزة أو الصم ولا تتطلب معرفة باللغة سواء أكانت مكتوبة أم منطوقة، وقد تنتمي إلى فئة الاختبارات غير اللفظية Nonverbal التي تتطلب معرفة باللغة المنطوقة ولكن لا تتطلب إجادة القراءة والكتابة كالاختبارات التي تتطلب فهم وتفسير الصور ورسوم المكعبات وغيرها.

ويمكن أن نضيف إلى شروط الإجراء الزمن المعطى للاختبار. فالاختبار إما أن يكون اختبار قوة وفي هذه الحالة تكون القدرة هي العامل الأساسي ويُعطى المفحوصون الوقت الكافي لإظهار أقصى ما لديهم من قدرة، وإما أن يكون اختبار سرعة وهنا يكون متغير السرعة بحد ذاته هو العامل الأساسي ويكون الغرض من الاختبار الكشف عن سرعة المفحوصين في الأداء الاختباري والتمييز بينهم تبعاً لذلك. ويضمّ اختبار السرعة عادة عدداً من الأسئلة اكبر من أن يجيب عنها المفحوصون في حدود الزمن المعطى للاختبار (بما في ذلك المفحوص الأسرع في المجموعة). وأما اختبارات القوة فتضمّ أسئلة متدرجة الصعوبة عادة وتستهدف

الكشف عن أعلى مستوى يمكن أن يصل إليه المفحوص، ويراعى فيها أن يكون الزمن كافياً للمفحوصين أو للأكثرية الساحقة منهم.

تصنيف الاختبارات انطلاقاً من مستوى التحديد في المهمات والإجابات:

من الاختبارات ما يضمّ أسئلة محددة وتنحصر مهمة المفحوص فيها بإعطاء إجابات محددة، ومنها ما يطرح أسئلة عامة وغير محددة وقد يتضمّن مثيرات غامضة بحد ذاتها وليس لها دلالة محددة، ويتطلب، بالتالي، إجابات حرة غير مقيدة نسبياً. إلى النوع الأول من الأدوات تنتمي الأدوات التي تندرج تحت اسم الاختبارات الموضوعية التي قد تتضمّن عدداً من العبارات الصحيحة والخاطئة وتنحصر مهمة المفحوص فيها بوضع إشارة أو كلمة صح أو خطأ كما في اختبارات صواب – خطأ، أو تتضمّن بنوداً أو أسئلة يُعطى لكل منها عدد من البدائل وتكون مهمة المفحوص هي اختيار البديل الصحيح بين تلك البدائل (اختبار الاختيار من متعدّد)، أو تتضمّن قائمة بالمقدمات أو المسائل وقائمة أخرى بالإجابات في ترتيب مختلف عن الأولى وتتطلب المزاوجة أو المطابقة بينهما (اختبار المطابقة). كما تنتمي إلى النوع الأول من الأدوات تلك الأدوات التي يطلق عليها اسم الاختبارات المحددة البناء حيث تكون المهمات محددة والإجابات مقيدة بالاحتمالات الواردة، وتتطلب الإجابة بـ "نعم" أو "لا" أو "لا أدري"، أو "موافق" أو "غير موافق"، أو "موافق جداً " – "موافق" – "محايد" – "معارض" – "معارض جداً " كما في استخبارات الاتجاهات والميول وبعض مقاييس الشخصية.

وتنتمي إلى النوع الثاني من الأدوات تلك الأدوات التي تعطي الحرية للمفحوص لتأليف الإجابة من عنده والتعبير بأسلوبه، كما تنتمي إلى هذا النوع تلك الأدوات التي تنطوي على مواقف أو مثيرات غامضة وليس لها معنى محدد بالأصل ولكن تمثّل مواقف من الحياة على المفحوص أن يعطيها معنى. ويطلق على هذه الاختبارات اسم الاختبارات محددة البناء جزئياً كاختبار تفهم الموضوع، بالإضافة إلى الاختبارات الإسقاطية غير المحددة في بنائها التي تضمّ

مثيرات غامضة ولا علاقة لها بالمواقف الحياتية كاختبار الرورشاخ لبقع الحبر. والغرض من عدم التحديد هو " إتاحة" الفرصة لذاتية الفرد وإسقاطاته وتنظيمه للموقف أو إدراكه له، وفي هذا ما فيه من نواحي الإسقاط والتعبير الحر إلى آخر الجوانب اللاشعورية (احمد، 1960، ص116).

أغراض الاختبارات النفسية:

يمكن إجمال الأغراض الأساسية التي تؤديها الاختبارات النفسية فيما يلي:

1. **التشخيص النفسي:**

لا شك أن محاولة الكشف عن حالات التخلف والضعف العقلي كانت من العوامل المهمة في ظهور اختبارات الذكاء وتطوّرها. إذ من المعلوم أن مقياس بينيه ظهر أساساً بهدف تعرّف أولئك التلاميذ الذين يعانون من تخلف عقلي يمنعهم من مواكبة زملائهم في الدراسة. بالإضافة إلى ذلك تتيح اختبارات الذكاء الكشف عن المتفوقين عقلياً أو الموهوبين، وعن المتوسطين أو الأسوياء بطبيعة الحال. واستناداً إلى رقم حاصل الذكاء ظهرت محاولات لتصنيف مستويات الأداء العقلي للأفراد بدءاً بأدنى تلك المستويات وانتهاءً بأعلاها من مثل التصنيف الذي اعتمد حاصل ذكاء ستانفورد – بينيه والتصنيف الذي اقترحه فكسلر كما سنرى ذلك بالتفصيل فيما بعد. ومن المفيد الإشارة هنا إلى أن رقم حاصل الذكاء اعتُمِد أساساً للتمييز بين ثلاثة مستويات أو ثلاث فئات للضعف العقلي هي:

1. المعتوهون Idiots: ويتميّزون بضعف عقلي شديد حيث يقل حاصل ذكائهم عن 25، ونسبتهم إلى المجموع الكلي للسكان هي 0.1٪. ويعجز المعتوهون عادة عن التواصل مع الآخرين عن طريق اللغة ويستخدمونها بصورة بدائية جداً، كما يحتاجون إلى من يحميهم من الأخطار الخارجية ويرعاهم حتى في الأمور المتصلة بإشباع حاجاتهم البسيطة.

2. البلهاء Imbeciles: ويتصفون بضعف عقلي متوسط الشدة حيث يتراوح حاصل ذكائهم من 25 إلى 50، ونسبتهم إلى المجموع الكلي للسكان تصل إلى 0.6 ٪. ويستطيع الأبله عادة حماية نفسه من الأخطار وتعلّم بعض الأعمال السهلة واستخدام الألفاظ على نطاق محدود ولكنه يعجز عن تعلّم القراءة والكتابة.
3. المورون Morones: وتتصف هذه الفئة بضعف عقلي خفيف حيث يتراوح حاصل ذكائها بين 50 و 75، ونسبة أفرادها إلى المجموع الكلي للسكان هي 1.3٪. ويستطيع أفراد هذه الفئة تعلّم المبادئ الأولى للقراءة والكتابة والحساب ويعجزون على الأغلب عن متابعة الدراسة بعد الصف الرابع الابتدائي.

ويميل أغلب العلماء في الوقت الحاضر إلى استخدام رقم حاصل الذكاء في التشخيص الأولي للضعف العقلي وعدم الاقتصار عليه بمفرده في إطلاق صفة الضعف العقلي على الطفل وبالتالي حرمانه من فرص التعليم وغيرها. ومن المؤكد أن الدراسة الشاملة للطفل التي تتناول مختلف جوانب نموه بما في ذلك نموه الجسمي والانفعالي والاجتماعي يمكن أن تفيد في تقديم صورة افضل عن الطفل وعن مستوى أدائه العقلي.

ولا يقتصر التشخيص النفسي على الكشف عن حالات التخلف والتفوق العقلي بل يشمل الاضطرابات النفسية بأنواعها والحالات المرضية ومظاهر الاستواء وعدم الاستواء. ومن خلال التحليل الكمي للدرجات التي يحصل عليها المفحوص في عدد من الاختبارات، والتحليل الكيفي لطريقة الإجابة ومضمونها، يمكن الخروج بعدد من الدلالات الإكلينيكية المهمة وتعيين مواطن الضعف في أداء المفحوص، والصعوبات التي يواجهها. والواقع إن التشخيص النفسي والدراسة التحليلية والمعمقة للحالات الفردية ترتكز على أدوات القياس والتقويم المختلفة بما في ذلك الأدوات التي تتصدّى لظواهر سوية أو سمات غير مرضية وتلك التي تتصدّى لقياس المواقف والميول. ولا نستطيع أن نقول: إن بعض الادوات ذات

استخدام إكلينيكي وبعضها الآخر بعيد عن هذا الاستخدام، "فكل اختبار يمكن عدّه اختباراً إكلينيكياً بالمعنى الحرفي للكلمة مادام يستخدم في مساعدة الأفراد وتحليل أدائهم بصورة أو بأخرى لمعرفة نواحي القوة والضعف فيه". (Freeman,1962,p.370). وقد أسهمت الممارسة الإكلينيكية في تطوير العديد من الاختبارات والمقاييس لمواجهة مشكلات التشخيص النفسي ومتطلباته.

وعموماً يمكن القول: إن الاختبارات النفسية بمجموعها هي مصدر ثرّ من مصادر المعلومات عن الفرد، وإن هذه المعلومات التي قد يتعذّر الحصول عليها بوسائل أخرى تفيد في تشخيص جوانب القوة والضعف سواء فيما يتصل بأدائه العقلي أم بتكيفه وسلوكه الاجتماعي وشخصيته ككل.

2. التشخيص التربوي:

ويستهدف هذا النوع من التشخيص الكشف عن صعوبات التعلّم التي يواجهها بعض الأفراد، ويتطلب استخدام الاختبارات النفسية بالإضافة لأدوات التقويم التربوي المختلفة. ويمكن تمييز ثلاث مراحل أساسية في عملية التشخيص التربوي هي:

1. تحديد أو تعيين التلاميذ الذين يواجهون صعوبات خاصة في التعلّم. وإحدى الطرائق المتبعة في ذلك مقارنة نتائج الاختبارات التحصيلية ولاسيما المقننة منها بنتائج اختبارات الذكاء والاستعداد المدرسي، فإذا كان مستوى التحصيل أدنى من مستوى الذكاء أو الاستعداد المدرسي لدى تلميذ معين (أو مجموعة من التلاميذ) فإن هذا يشير إلى وجود صعوبات دراسية يواجهها هذا التلميذ (أو التلاميذ). ومن الطرائق المتبعة أيضاً دراسة وتحليل الصفحة النفسية للتلميذ "البروفيل" التي تضمّ عادة النتائج التي يحصل عليها في مجموعة كبيرة من الاختبارات أو "بطارية" اختبارات تغطي مجالات دراسية عديدة. ويفيد هذا التحليل في مقارنة

إنجاز التلميذ في كل مجال بمستوى إنجازه العام، فإذا ظهر ضعف في مجال ما أو مهارات معينة فهو يدل على الصعوبة (أو الصعوبات) التي يعاني منها التلميذ.

2. تحديد الطبيعة الخاصة للصعوبة ومواطن القوة عند التلميذ. فصعوبات التعلّم على درجات، وفي بعض الحالات يمكن الاكتفاء بالمعلومات التي تقدّمها الإجراءات العامة السابقة والانتقال منها مباشرة إلى العمل العلاجي. وفي حالات أخرى قد نحتاج إلى المزيد من الدراسة التشخيصية قبل التخطيط للعمل العلاجي وهذا يتطلب استخدام اختبار تشخيصي وتحليل استجابة التلميذ لكل بند من بنوده ولاسيما إذا ارتبطت الصعوبة بإحدى المهارات الأساسية في القراءة والكتابة والحساب. ومن المفيد في هذه المرحلة من مراحل عملية التشخيص العمل على تحديد نقاط القوة لدى التلميذ إذ يمكن اعتماداً عليها مواجهة الصعوبة وتجاوز الضعف. فالعلاج الفعال لنقاط الضعف يتطلب الاعتماد على جوانب القوة، وينطلق من أن التلميذ يجب أن يتحسس نجاحاته. فإذا اعتمد على جوانب قوته ازدادت فرص النجاح أمامه وازدادت أمامه بالتالي فرص التغلب على الصعوبات (Ahmann, 1975).

3. تحديد عوامل الضعف. فقد تعود بعض صعوبات التعلّم إلى طرائق التعليم أو المادة التعليمية شديدة الصعوبة. وهذا النوع من الصعوبات يمكن الكشف عنه بسهولة ولاسيما عندما يواجه عدد كبير من التلاميذ الصعوبة نفسها. غير أن الكثير من الصعوبات الدراسية يمكن أن تحدث نتيجة عوامل أخرى بينها الحالة الصحية، والبيئة المنزلية، والمصاعب التكيفية، وعادات الدراسة. بالإضافة إلى مستوى النمو العقلي العام للمتعلّم وقدراته الخاصة وميوله مما يظهر أهمية الاختبارات النفسية في تشخيص تلك الصعوبات.

ولابد من الإشارة هَهُنا إلى أن التشخيص التربوي بمفهومه الحديث لا يقتصر على المعارف والمهارات الأكاديمية، فقد اتّسع مجاله الآن لينسجم مع

المفهوم الحديث للتربية التي تلحّ على مظاهر النمو كافة. وقد أظهرت البحوث أن العديد من العوامل والمتغيرات " غير المعرفية" من مثل التكيف الشخصي والاجتماعي للتلميذ، وشخصيته بصفة عامة، ونموه الانفعالي، لها أثرها المباشر في نموه المعرفي وتحصيله الدراسي. ويمكن القول إن التشخيص التربوي الحديث بدأ يتداخل إلى حد بعيد مع التشخيص النفسي ويتضافر معه في تقديم صورة شاملة ومتكاملة لشخصية التلميذ. وهذا كله يعزّز مكانة أدوات القياس والتقويم ويبرز استخداماتها التشخيصية الواسعة.

3. التوجيه والإرشاد:

يحتاج الدارسون إلى التوجيه خلال حياتهم الدراسية وفي اختيارهم المهني. ويهتم التوجيه بتعريف الطالب بقدراته واتجاهاته وميوله وبما يصلح له من دراسات ومهن بهدف مساعدته على الاختيار، واتخاذ قرارات صائبة حول مستقبله الدراسي والمهني. وللاختيار الدراسي والمهني أهميته البالغة ومنعكساته الخطيرة في حياة الفرد، وتظهر آثاره مباشرة في إنجازه ونجاحاته، (أو فشله) الدراسي والمهني، وفي تكيفه الشخصي والاجتماعي وشعوره بالرضا والسعادة أو تعرضه لخبرات الفشل والإحباط. فإذا أخذنا في الحسبان أن الآباء كثيراً ما يزجّون بأبنائهم في دراسات لا تتلاءم مع قدراتهم وميولهم برزت أهمية التوجيه الدراسي والمهني وضرورة استناده إلى أسس علمية وسليمة.

يتطلب التوجيه الدراسي والمهني معرفة موضوعية وواسعة بقدرات الفرد واتجاهاته وميوله وتكوين صورة متكاملة وصادقة عنه. وكلما كانت معرفتنا للفرد صادقة وشاملة واستطعنا من خلالها تبيّن جوانب القوة والضعف عنده ازدادت فعالية التوجيه وازداد احتمال تقديم المعونة المفيدة له ومساعدته على اختيار أفضل فرص التعلّم، ثم فرص العمل. وتمدنّا أدوات القياس والتقويم المختلفة كاختبارات التحصيل ومقاييس الذكاء والقدرات واستخبارات الميول والاتجاهات، بالإضافة إلى الملاحظة اليومية للمعلم، بمعلومات مهمة عن الفرد

يمكن استخدامها في مساعدته على تكوين صورة واضحة وواقعية عن نفسه وإمكاناته، ومساعدته بالتالي على حسن الاختيار. ويمكن استخدام هذه الأدوات في تخطيط المستقبل التعليمي للفرد من خلال تحديد المجالات التي يحتمل أن يتفوق فيها أكثر من سواها، والكشف عن "القدرات" الخاصة، والحثّ على صقلها وتطويرها.

تشير آنا أنسازي إلى أن استعمال الاختبارات اتسع مجاله للغاية وتجاوز نطاق عملية التوجيه الدراسي والمهني ليكون أساساً في عملية الإرشاد بالمعنى الشامل ويغطي مختلف مظاهر حياة الفرد. فالاستقرار العاطفي والعلاقات السليمة مع الأفراد أصبحت أهم أهداف الإرشاد. وثمة إلحاح متزايد على استعمال الاختبارات للمساعدة في فهم الذات ونمو الشخصية (Anastasi,1982). وقد أسهمت الاختبارات إسهاماً كبيراً في تطوّر حركة الإرشاد النفسي من خلال دورها في تنمية معارف الفرد عن نفسه وقدراته ومساعدته على " حسن الاختيار "، ومواجهة مشكلاته التكيفية الشخصية والاجتماعية. فإذا أضفنا إلى ذلك الاستخدامات التشخيصية الواسعة للاختبارات وأخذنا في الحسبان أن التشخيص النفسي والتربوي هو دعامة أساسية في عملية التوجيه والإرشاد، وأن الفصل بين عملية التشخيص وعملية التوجيه والإرشاد هو فصل اصطناعي وقد تسوّغه أغراض الدراسة فقط، برزت أمامنا مكانة التقويم وأدواته في عملية التوجيه والإرشاد بصورة واضحة وجلية.

4. الانتقاء والتصنيف وتحديد المسار التعليمي للدارسين:

تعتمد الإدارة التعليمية على أدوات القياس والتقويم المختلفة في اتخاذ جملة من القرارات المهمة المتعلقة بانتقاء الدارسين Selection وتصنيفهم Classification وتحديد مسارهم التعليمي أو وضعهم في المكان المناسب Placement. وتتمثّل عملية الانتقاء في قبول بعضهم لدراسة أو تخصص معين ورفض بعضهم الآخر من الراغبين في هذه الدراسة أو التخصص استناداً إلى نتائج

القياس والتقويم وفي حدود الأهداف المرسومة للمؤسسة التربوية المعنية. والتصنيف هو خطوة لاحقة بالانتقاء ومتممة له ويقوم على فرز الطلاب إلى مستويات وتحديد الضعاف والمتوسطين والمتفوقين بينهم بهدف "غربلتهم" فيما بعد واختيار الأقوى بينهم لمتابعة الدراسة. ويتجسد التصنيف في تقسيم الطلاب وتشعيبهم أي توزيعهم إلى شعب متجانسة أو متكافئة وذلك حسب نتائجهم في اختبارات التحصيل، أو اختبارات التحصيل والذكاء معاً، كما حدث بخاصة في مدارس الولايات المتحدة وإنكلترا لفترة طويلة من الزمن. وقد يتمّ التصنيف بتقسيم التلاميذ وتوزيعهم إلى أنواع التعليم المختلفة كالتعليم العام أو الفني أو الزراعي أو التجاري وما إلى ذلك. ويتبع التصنيف تحديد المسار التعليمي للتلميذ ووضعه في المكان "المناسب" أو المساق الدراسي المناسب استناداً إلى مستوى استعداده وتحصيله العام، أو تحصيله في مجال دراسي محدد كاللغة أو الرياضيات مثلاً.

أكدت التربية في العالم أجمع ولفترة طويلة من الزمن وظيفتها "الاصطفائية"، وألحّت على أغراض القياس والتقويم في الانتقاء والتصنيف وأعطتها الأولوية بين وظائفه وأغراضه انطلاقاً من أن القلة من الدارسين هم المؤهلون لمتابعة الدراسة والالتحاق بالجامعات. وكانت عملية انتقاء الدارسين وتصنيفهم إلى فئات متنوعة تتمّ استناداً إلى مستوياتهم كما تكشف عنها اختبارات التحصيل والقدرة في غالب الأحيان. وكثيراً ما استخدمت اختبارات الذكاء بخاصة بغرض الكشف عن قدرات التلميذ واستعدادته وقابليته للتعلّم في مرحلة مبكرة وقبل بلوغه الثانية عشرة من عمره مما أدى إلى الحدّ من فرص التعليم وحصرها بالقلة من الدارسين. وقد كان لهذه النظرة الاصطفائية مضارها وانعكاساتها الخطيرة على تعليم الأجيال. وكان من الطبيعي أن تهرم هذه النظرة على مر الزمن بتأثير عوامل ومتغيرات عديدة وأن تظهر إلى الوجود نظرة جديدة تدعو إلى توسيع فرص التعليم بصورة هائلة والوصول بالمتعلّم إلى أقصى ما تسمح به طاقاته وقدراته. وقد انتشرت هذه النظرة الجديدة ولقيت

قبولاً واسعاً في مختلف أرجاء العالم. ويتوقع أن تتعزّز هذه النظرة وتصبح حقيقة واقعة مع الدخول في عصر التعليم الجماهيري وتنامي الدعوة إلى الإفادة القصوى من الطاقات البشرية و"توظيفها" واستثمارها على النحو الأمثل.

ولا تلغي النظرة الجديدة إلى التربية عملية الانتقاء والتصنيف وتحديد المسار التعليمي للمتعلّم على الرغم من إلحاحها على نشر التعليم وتعميمه ودعوتها في الوقت الحاضر إلى تأمين فرص التعليم الأساسي (حتى انتهاء المرحلة الثانوية) للجميع ممن هم في سن الدراسة. والواقع أن النظام التعليمي، أي نظام تعليمي، سواء اتجه إلى التشدّد بدرجة ما وإتاحة الفرصة لأعداد محدودة في الدراسة التي يرغبون فيها (كما هو الحال في كليات الطب أو مدارس الطيران مثلاً)، أم اتجه إلى التوسّع والتساهل، لابد أن يلجأ إلى الانتقاء. والمنطلق في عملية الانتقاء أو الاختيار أنه يستحيل تلبية رغبات أي فرد لأية دراسة أو تخصص يريده ولاسيما في مرحلة مابعد التعليم الأساسي والمستويات الدراسية العليا. ومن الواضح أن عملية الانتقاء بمنظورها الجديد لابد أن تستند إلى أسس وقواعد محددة بمعنى أنه لابد من وجود نظام للانتقاء. والغرض من هذا النظام هو قبول الأشخاص الذين تتزايد أمامهم فرص النجاح في الدراسة المقبلة ورفض أولئك الذين تتضاءل أو تنعدم أمامهم تلك الفرص. وقد ثبت أن الأداء الراهن للفرد كما تقيسه اختبارات التحصيل والقدرة هو متنبئ صادق بدرجة عالية بأدائه المقبل في مجالات دراسية محددة. كما أن علامات المدرسة الثانوية هي بين أفضل عوامل التنبؤ بالتحصيل الجامعي في المستقبل.

(ثورندايك وهيجن، 1989، ص178)

وما من شك في أن التصنيف (أو التشعيب) بصورته الكلاسيكية لم يعد مقبولاً اليوم، وقد تبيّن من خلال العديد من الدراسات أنه ليس في مصلحة الضعاف والمتوسطين من الدارسين الذين يشكّلون الأكثرية بطبيعة الحال. والواقع أن التصنيف أو التشعيب القائم على توزيع التلاميذ في مرحلة مبكرة من

مراحل التعليم إلى مجموعات متجانسة استناداً إلى حاصل الذكاء والذي يؤدي إلى إتاحة الفرصة لنسبة ضئيلة منهم لمتابعة الدراسة أصبح مرفوضاً تماماً اليوم. وقد أكّد ليندكويست أنه يجب الايصنّف التلاميذ في مرحلة الدراسة الأولى تبعاً لأي أساس تحصيلي أو عقلي، واقترح أن ينقل التلاميذ بصورة آلية. كما أكدّ أن تكييف عملية التعلّم للفروق الفردية وحاجات التلاميذ يقتضي الا نضع لجميع التلاميذ مستوى واحداً يطلب منهم الوصول إليه للحصول على شهادة التعليم العام، وإنما يجب أن تكون هذه الشهادة هي مجرد شهادة بانقضاء مدة معينة (الغريب، 1970). وفي هذه الحالة يستخدم القياس بهدف مقارنة كل تلميذ بنفسه وبما لديه من استعدادات، كما يقارن بغيره ممن لهم استعداده نفسه ومن قضوا مدة التعليم نفسها. وتتّجه التربية الحديثة إلى القبول بمبدأ التصنيف من منطلق الاستجابة للفروق الفردية بين الدارسين فقط، بحيث يؤدي هذا التصنيف إلى تسهيل عملية التعلّم والاستمرار بها إلى الحدود القصوى التي تسمح بها قدرات التلميذ بدلاً من إضاعة الفرصة على أغلبية التلاميذ و"دمغ" بعضهم مسبقاً بالتخلف والقصور وانطلاقاً من هذه النظرة ذاتها تقوم عملية توجيه المسار التعليمي للفرد على تقديم البدائل الملائمة لهذا الفرد وإتاحة الفرصة له ليتحرك حسب سرعته الخاصة وبما يتلاءم مع مستوى المعارف والمهارات التي يمتلكها في كل من المجالات الدراسية المختلفة. ويمكن من خلال هذه العملية اتخاذ قرار حول ما إذا كان الطالب (س) سيلتحق بالشعبة العامة في دروس الرياضيات أم أنه يحتاج إلى التدريس العلاجي في هذه المادة أم أنه، على العكس، يمكن أن يدخل إلى مساق متقدّم في هذه المادة.

تؤدي الاختبارات النفسية دوراً مهماً للإدارة التعليمية التي تتولى عملية الانتقاء والتصنيف وتوجيه المسار التعليمي بمنظورها الجديد. ومن أجل الوصول إلى قرارات سليمة بهذا الشأن يحتاج صانع القرار إلى معلومات صادقة ودقيقة لكي تساعده على التنبؤ بأداء الفرد وقدرته على التعلّم، مما يتطلب الاعتماد على أدوات عالية الجودة لقياس الاستعداد والتحصيل والميول. ولعل من بين الأمور التي

لا يصحّ إغفالها بأي حال من الأحوال هو أن عملية الانتقاء والتصنيف تفرض شروطاً صارمة في أدوات القياس بينها ما يتصل بتوظيفها بالاتجاه المناسب وعدم إساءة استعمالها، وبينها ما يتصل بقدرتها التنبؤية والتشخيصية العالية بالإضافة إلى مواصفاتها الفنية العامة وحساسيتها للفروق الدقيقة.

5. البحث العلمي في المجالات التربوية والنفسية:

تشغل الاختبارات النفسية حيزاً مهماً ضمن فعاليات البحث وتعدّ الركيزة الأساسية في عملية البحث في الكثير من مراحلها وخطواتها. ومع أن الغرض الأول والأهم للبحث هو تكوين المعرفة حول الظاهرة موضع الدراسة أو الاستزادة منها بينما تنصب عملية القياس والتقويم، على الأغلب، على تلبية أغراض عملية مباشرة وتسعى إلى الحكم على فاعلية أو قيمة ما نريد قياسه وتقويمه، فإن ثمة نقاطاً مشتركة بين عملية القياس والتقويم وعملية البحث تؤدي إلى تداخلهما بل والخلط بينهما في حالات. بين هذه النقاط ما يتصل بالمنهجية الصارمة وشروط الموضوعية والحياد والدقة التي تتطلبها كل من هاتين العمليتين، وبينها ما يتصل بالأدوات المستخدمة والإجراءات المتبعة في كل منهما، هذا بالإضافة إلى التداخل في الأغراض التي تسعى كل منهما إلى تلبيتها في الكثير من الحالات. ومن الواضح أنه بناءً على نتائج البحوث العلمية يمكن اتخاذ العديد من القرارات التي تمس العملية التربوية بمختلف جوانبها كالقرارات الإدارية والتعليمية وغيرها.

تعتمد البحوث النفسية والتربوية بأنواعها على أدوات القياس والتقويم لجمع المعلومات والبيانات، ويعتمد بعضها، كالبحوث التجريبية على تلك الأدوات للتحقق من صحة الفرضيات المطروحة. وغني عن البيان أن البحوث التي أجريت في نطاق علم نفس الفروق الفردية بأنواعها ولاسيما بحوث الوراثة اعتمدت بصورة أساسية على اختبارات الذكاء. والواقع أن أدوات القياس والتقويم المختلفة يمكن عدّها جميعاً أدوات بحث حيثما تمّ توظيفها لأغراض البحث

وأمكن من خلالها الحصول على معلومات يحتاج إليها الباحث، بغض النظر عن الأغراض الأخرى التي تسعى إلى تلبيتها أساساً كالأغراض التشخيصية مثلاً. وكثيراً ما يعمد الباحث إلى الإفادة من البيانات والمعلومات التي تقدّمها تلك الأدوات في اشتقاق الفرضيات بالإضافة إلى التحقق من الفرضيات، مما يدل على المكانة المهمة لتلك الأدوات في عملية البحث.

ومن المفيد الإشارة هَهُنا إلى أن أداة القياس التي تستخدم كأداة بحث وسواء حصل عليها الباحث بصورة جاهزة أم عمد إلى تصميمها بنفسه، تتطلب مراعاة جملة من الشروط والمواصفات الفنية المهمة بينها ما يتصل بصلاحيتها وتوفير قدر عالٍ من صدقها وموثوقيتها (ثباتها)، وبينها ما يتصل بعملية إجرائها وتطبيقها واستثارة الدافعية لدى المفحوصين "لأخذها" والإجابة عنها. فإذا أخذنا في الحسبان أن متغير الدافعية هو أحد المتغيرات المهمة في الموقف الاختباري، وأن دافعية المفحوصين للأداء الاختباري كثيراً ما تضعف في المواقف الاختبارية التي يتمّ إعدادها أو تصميمها لأغراض البحث ظهرت أمامنا ضرورة العمل على رفع مستوى الدافعية لدى المفحوصين وإتاحة الفرصة لهم للتعبير عن أدائهم الحقيقي. ويستتبع هذا بطبيعة الحال "تقنين" الأداة وتوحيد سائر الشروط والعوامل المؤثرة في الأداء الاختباري مع ملاحظة أن تقنين الأداة لأغراض البحث لا يتطلب استخراج أو اشتقاق معايير للأداء إلا إذا كان الغرض من البحث ذاته هو استخلاص تلك المعايير.

الفصل الثاني

تصميم الاختبار النفسي وتقنينه



الفصل الثاني
تصميم الاختبار النفسي وتقنينه

تمثّل عملية تصميم الاختبار وتقنينه مهمة بالغة الصعوبة، كما تنطوي على قدر كبير من الأهمية، وتعكس بشكل أو بآخر شخصية المرء الذي يتصدّى لها وبراعته و"إبداعه" الخاص. وعلى الرغم من أن هذه العملية تخضع لمجموعة من الأسس والقواعد المنهجية وتتطلب استخدام جملة من الأساليب والإجراءات المحددة التي توصلت إليها نظرية القياس المعاصرة وممارساته المتقدّمة فإن العنصر "الفني" أو الإبداعي في هذه العملية يصعب التقليل من شأنه ومكانته. ولا يتسع المقام في الفصل الحالي لتقديم دراسة تفصيلية ومعمّقة لعملية تصميم الاختبار وتقنينه بكل ما تتضمّنه هذه العملية من خطوات وإجراءات تنفيذية وما تتطلبه من شروط وقواعد صارمة. وكل ما نطمح إليه في هذا الفصل هو إلقاء الضوء على هذه العملية والتعريف بخطواتها الأساسية ودورها الحاسم في إمدادنا باختبارات نفسية تتمّتع بمواصفات جيدة.

أسس تصميم الاختبار النفسي:

تنطلق عملية تصميم الاختبار من مجموعة من الأسس والمبادئ وتخضع لخطة عمل محكمة، تتحدّد فيها مسبقاً جملة المراحل والخطوات والإجراءات التنفيذية اللازمة. وسيكون من المفيد الوقوف أولاً عند الأسس والمبادئ العامة المعتمدة في تصميم الاختبار النفسي التي تحدّد في حقيقة الأمر جميع الخطوات والإجراءات اللاحقة.

ومن الأسس المهمة في تصميم الاختبارات النفسية ما يلي:

1. إن عملية تصميم الاختبار يجب أن تأخذ بالحسبان الهدف أو الأهداف الرئيسة للاختبار، إذ قلما يعدّ تصميم اختبار جديد وإعداده هدفاً قائماً بذاته، بل وربما لا يعدّ كذلك إطلاقاً. والباحث عادة لا يعمد إلى تصميم اختبار جديد إلا بعد الاطلاع على الاختبارات المتوفرة والتأكد من أنها لا تفي بالغرض المطلوب، أو أنها تلبي هذا الغرض ولكن بدرجة غير كافية. ومع أن الاختبار النفسي بالأساس هو أداة علمية للبحث والقياس تنطوي على قدر كبير من الأهمية على الصعيد العملي المحض، وأنّ تصميم هذه الأداة جدير بالتالي أن يكون هدفاً بحد ذاته فإن الاختبارات النفسية ظهرت في حقيقة الأمر لتلبية حاجات عملية مباشرة أملتها أوضاع وشروط ومطالب واقعية محددة. ولا يستثنى من ذلك تلك الاختبارات التي ارتكزت مباشرة إلى أسس نظرية معينة بل وسعت إلى التثبت من فرضيات علمية محددة، كما أسهمت إسهاماً كبيراً في إغناء وتطوير معارفنا النظرية عن تلك السمات أو الخصائص التي سعت إلى قياسها من مثل اختبار بينيه وسيمون الشهير للذكاء الذي استهدف أساساً كما هو معلوم فصل أولئك الأطفال الذين يعجزون عن متابعة زملائهم في الدراسة وتقديم المعونة الخاصة لهم.

2. إن تحديد الهدف من الاختبار لابد ان يتبعه (أو يرافقه) تحديد الغرض الخاص له (التشخيص أو الفرز أو التصفية السريعة مثلاً)، ومجالات استعماله، والمجتمع الأصلي للأفراد الذين سيطبق عليهم، ومن هم هؤلاء الأفراد؟ وما مستواهم التعليمي وسنهم؟.. إلخ. كما لابد من تحديد نوع هذا الاختبار كأن يكون لفظياً او أدائياً أو لفظياً وأدائياً معاً، وطريقة إجرائه كأن يطبق فردياً أو جميعاً وغير ذلك من الأمور التي تتصل مباشرة بالفكرة الموجّهة لعملية تصميم الاختبار وتؤدي بالمحصلة النهائية إلى تلبية الهدف المرسوم له.

3. يجب وضع البنود الاختبارية بحيث تتيح إنتاج أو استدعاء عينة من الاستجابات تكون ممثلة لمجتمع السلوك المراد قياسه وتغطّي مظاهر هذا السلوك وجوانبه المختلفة مع إعطاء الوزن النسبي لكل منها. فالاختبار النفسي بالأصل هو عينة من البنود تؤخذ من المجتمع الأصلي من البنود التي كان يمكن استخدامها في الاختبار. كما أن استجابات أو إجابات المفحوصين عن تلك البنود (أو المثيرات) يفترض أن تكون عينة صادقة وممثلة لاستجاباتهم المحتملة عن المثيرات أو البنود كلّها التي كان يمكن استخدامها في هذا الاختبار. وبعبارة أخرى فإن أداء المفحوصين في الاختبار يفترض أن يعكس أداءهم في ذلك المجال "الأوسع" الذي يتصدّى له الاختبار ويسعى إلى تغطيته.

وكلما كان مجال القدرة أو السمة التي يسعى الاختبار إلى تغطيتها كبيراً كان المجتمع الأصلي للبنود كبيراً وقلّ تجانسه على الأرجح، مما يتطلب زيادة عدد بنود الاختبار للحصول على عينة أصدق وأقدر على تمثيل المجتمع الأصلي بطبقاته وفئاته المختلفة مع مراعاة الوزن النسبي لكل منها. وحيث أن الاختبار، أي اختبار، هو عينة سحبت من مجتمع أصلي ما فإن درجة المفحوص في هذا الاختبار تتضمّن قدراً من خطأ العينة. وبطبيعة الحال كلما كانت العينة أكبر أعطت مؤشرات إحصائية أقرب إلى مؤشرات المجتمع الأصلي وأصبح الخطأ أقلّ.

4. يجب وضع البنود الاختبارية بحيث لا تؤثّر العوامل الجانبية أو الدخيلة في أداء المفحوص وتمنعه عن إظهار هذا الأداء على حقيقته. ومن هذه الزاوية يجب الابتعاد عن الجمل الطويلة المعقدة والمفردات الصعبة أو التي تحمل أكثر من معنى ما لم تكن هذه الجمل والمفردات ذاتها موضوعاً للقياس، كما لابد من تجنب التلميحات والإيحاءات وكل ما يمنع الاختبار من العمل بالاتجاه المطلوب والكشف عن الفروق الحقيقية في أداء المفحوصين.

5. يجب أن تكون البنود من مستوى صعوبة ملائم وأن تتمتّع بقدرة تمييزية عالية. ومن المعلوم أن المستوى الملائم لصعوبة البنود في اختبارات القدرة هو 50٪ وأن البنود التي تصل معاملات سهولتها (أو صعوبتها) إلى 50٪ أو تقترب من ذلك بحيث تتراوح من 40٪ إلى 60٪، هي اقدر البنود على التمييز أو التفريق بين المفحوصين في السمة أو القدرة المقيسة. وتبعاً لذلك لابد أن تخضع البنود للتحليل الإحصائي ليصار إلى تنقيتها و "غربلتها" واستبعاد السهلة جداً أو الصعبة جداً منها في ضوء نتائج هذا التحليل. وغني عن البيان أن التحليل الإحصائي للبنود لا يجوز أن يحل محل التحليل المنطقي الذي يسبق هذا التحليل ويرافقه أيضاً، أو يكون بديلاً عنه. والأصح أن يكون التحليل الإحصائي متمّماً للتحليل المنطقي وأن يتضافر هذان النوعان من التحليل للوصول إلى اختبار عالي الجودة.
6. لابد من توافر شرط التقنين في الاختبار النفسي. ويتضمّن التقنين توحيد إجراءات التطبيق، كما يتضمّن توحيد عملية تفسير دلالات الدرجات الخام من خلال المعايير التي يزوّد بها الاختبار. هذا بالإضافة إلى توفير مستلزمات الصدق والثبات كما سنرى.
7. ان ترتّب بنود الاختبار وفق مبدأ التدرج في الصعوبة بحيث تبدأ بأسهل البنود وتنتهي بالأصعب منها. فإذا رتبت البنود على هذا النحو فإن اي مفحوص سيجيب عن عدد معين منها ويقف عند حد معين لا يستطيع تجاوزه يقابل مستوى الصعوبة الملائم له. كما أن ترتيب البنود على هذا النحو سيرفع مستوى الدافعية لدى المفحوص وسيمنع احتمال ضياع الكثير من وقت المفحوص وجهده في محاولة الإجابة عن بنود يتعذّر عليه الإجابة عنها.

خطوات تصميم الاختبار النفسي:

تظهر بين الباحثين اختلافات غير ضئيلة أحياناً حول خطوات تصميم الاختبار وتتابعها. غير أن أغلب الباحثين يميلون إلى تأكيد الخطوات الأساسية التالية:

1) تحديد الهدف العام للاختبار والفكرة الموجِّهة له:

فتحديد وبلورة الهدف العام للاختبار أو الفكرة الموجهة له ينطوي على تصور أولي للاختبار يمكن أن يتمّ في ضوئه وضع الخطوط العريضة لخطة بناء الاختبار والتنبّه مسبقاً للاحتمالات والشروط المساعدة أو "المعرقلة" لعملية تطوير الاختبار وإخراجه إلى حيز الوجود.

2) تحديد الغرض أو الأغراض الخاصة للاختبار:

ينطوي تحديد الغرض أو الأغراض الخاصة للاختبار على قدر كبير من الأهمية نظراً لأنه يعدّ بمثابة ترجمة للفكرة الموجهة للاختبار أو الهدف العام له، والذي قد يكون غائماً بحد ذاته أو غير واضح بدرجة كافية، إلى أغراض واضحة ومحددة، كما يمثّل نقطة البداية الفعلية في بناء الاختبار ويؤثر في الخطوات اللاحقة جميعاً. فتحديد الغرض أو الأغراض الخاصة للاختبار يتضمّن مزيداً من التفصيل حول ميدان استخدامه والاستعمال الخاص له مما ينعكس مباشرة على إجراءات تصميمه. فإذا كان الغرض من الاختبار الكشف عن الفروق الدقيقة بين الأفراد والحصول على بيانات صادقة ودقيقة عنهم في السمة أو القدرة المقيسة كان من الضروري استخدام الطرائق والأساليب الإحصائية والفنية الراقية التي تلبّي هذا الغرض وتتيح التمييز بينهم استناداً إلى مستويات القدرة لديهم. وإذا كان الغرض من الاختبار الكشف عن الصعوبات (أو الاضطرابات) التي يعاني منها الأفراد وتشخيص جوانب القوة والضعف في أدائهم، كان لابد أن يتضمّن الاختبار عدداً كبيراً من البنود لتغطية المجال المراد

تشخيصه وأن تتيح هذه البنود تعرّف الصعوبات التي يعاني منها المفحوصون وتحديدها، بالإضافة إلى تحديد نقاط القوة التي يمكن أن تفيد في مواجهة تلك الصعوبات. أما إذا كان الغرض من الاختبار هو جمع بيانات سريعة أو إجراء مسح سريع بهدف التصفية أو الغربلة لأعداد كبيرة من المفحوصين كان بالإمكان التساهل قليلاً في بعض شروط الإجراء والاستعانة بأشخاص غير مدربين تدريباً عالياً في تطبيق الاختبار، هذا بالإضافة إلى استخدام معايير بسيطة وسهلة لتفسير نتائج الاختبار.

3) تحديد المجال الخاص الذي يتصدّى الاختبار لقياسه وعينة السلوك الممثلة له:

فإذا كان الاختبار موجهاً لقياس سمة الذكاء أو الانطواء أو القدرة العددية مثلاً كان من الضروري تحديد المجال الخاص لهذه السمة أو القدرة وما تتضمّنه من " عناصر " أو "مكونات" رئيسة وفرعية مع تحديد الأوزان النسبية لكل منها لكي يصار إلى تخصيص عدد من البنود لكل عنصر استناداً إلى زنه النسبي. ومن الواضح أن تحديد المجال الخاص للسمة المقيسة أو موضوع القياس يمثّل مهمة من المهمات الصعبة نظراً لأن السمات النفسية من مثل الذكاء والقدرة العددية والانطوائية وغيرها تنطوي على قدر كبير من العمومية والتجريد والتعقيد. وتفيد التعريفات الإجرائية لتلك السمات، بلا شك، في ترجمتها إلى وقائع سلوكية واستجابات ظاهرة مما يتيح إخضاعها للقياس من خلال التصدّي لعينة منها يفترض أن تمثّلها بجوانبها ومكوناتها المختلفة.

ومن نافلة القول: إن تحديد المجال الخاص الذي يتصدّى الاختبار لقياسه وعينة السلوك الممثلة له لابد أن يتبعه (أو يرافقه) تحديد المجتمع الأصلي للأفراد الذين سيطبق الاختبار عليهم ليصار فيما بعد إلى تطبيق الاختبار بهدف تجريبه وتعييره على عينات عشوائية من المجتمع الأصلي ذاته الذي صمّم له الاختبار. فإذا صمّم الاختبار لقياس ذكاء المرشحين لدخول

الجامعة سحبت العينات من أولئك الأفراد الذين تقدّموا بطلبات التحاق إلى الجامعة، وإذا صمّم الاختبار لقياس ذكاء الصم وبعض ذوي الإعاقة في المدى العمري الذي يتراوح من 6 إلى 12 سنة مثلاً سحبت العينات من تلك الفئة من الأشخاص حصراً وهكذا....

4) تحديد زمن الاختبار وطوله:

فمصمّم الاختبار لابد أن يحدّد مسبقاً، وقبل أن يبدأ بإعداد البنود، الزمن الذي سيعطى للمفحوصين للإجابة عن بنود الاختبار وعدد هذه البنود. وما من شك في أن زمن الاختبار وطوله يتحدّدان بالأغراض الخاصة له. فإذا كان الاختبار يهدف إلى المسح والمقارنة السريعة أو التصفية وسيطبق على أعداد كبيرة من المفحوصين تضاءلت الحاجة إلى زيادة عدد بنوده، وأما إذا كان الغرض من الاختبار هو تصنيف المفحوصين استناداً إلى مستويات القدرة لديهم واتخاذ قرارات مهمة حول مستقبلهم الدراسي أو المهني تزايدت الحاجة إلى زيادة عدد البنود ورفع القدرة التمييزية لهذه البنود. وفي الحالات التي يكون الغرض من الاختبار فيها تشخيصياً تصبح الحاجة ماسّة لزيادة عدد البنود إلى الحدود القصوى المتاحة وتغطية ذلك المجال الخاص والضيق الذي يتصدّى له الاختبار بهدف تعرّف الصعوبات ومواطن الضعف والقوة في أداء المفحوص دون إيلاء أهمية كبيرة للقدرة التمييزية للبنود. ومن الأمور التي يجب أخذها بالحسبان عند تحديد زمن الاختبار وطوله الشكل أو الأشكال التي ستأخذها البنود. فالأسئلة التي تتطلب من التلميذ أن يضع أو يؤلف الجواب بنفسه كأسئلة التكميل أو الأسئلة الإنشائية تحتاج إلى وقت أطول من الأسئلة الموضوعية التي تتضمّن عادة عدداً من البدائل ولا تتطلب سوى وضع إشارة أو رقم عند البديل الصحيح. كما أن البنود التي تتطلب أداء عملياً (البنود غير اللفظية) تحتاج إلى وقت أطول على الأغلب من البنود اللفظية. وبطبيعة الحال فإن لعدد البنود أهميته في رفع مستوى ثبات الاختبار وصدقه كما سنرى

5) إعداد البنود الاختبارية بصورتها الأولى:

وتعدّ هذه الخطوة الخطوة الحاسمة والأهم في تصميم الاختبار. حيث تظهر هنا مهارة مصمّم الاختبار وبراعته وقدرته "الإبداعية" الخاصة على تطوير وتأليف بنود اختبارية تلائم شكلاً ومضموناً الغرض أو الأغراض التي يسعى الاختبار إلى تلبيتها وتكون بمثابة عينة ممثلة لمجال السلوك المراد قياسه. وعلى الرغم من أن ثمة عدداً من القواعد المنهجية والأسس والتوصيات التي تساعد، في حال العمل بموجبها، في الوصول إلى اختبار جيد فإن تلك القواعد والأسس والتوصيات لا تضمن بحد ذاتها الوصول إلى مثل هذا الاختبار. والحق إنّ عملية تطوير وإعداد البنود الاختبارية تعتمد على شخصية مصمّم الاختبار بالدرجة الأولى. غير أن هذا لا يقلّل بطبيعة الحال، ولا يجوز أن يقلّل، من أهمية التحليل الكيفي والكمي لتلك البنود والذي يتعين على مصمّم الاختبار أن يجريه تمشياً مع جملة القواعد المنهجية والأسس والتوصيات التي توصلت إليها نظرية القياس الحديثة وممارساته المتقدّمة. وغني عن البيان أن العدد الكلي للبنود في هذه المرحلة لابد أن يزيد زيادة كبيرة على العدد الكلي للبنود التي سيضمّها الاختبار بصورته الأخيرة. فإذا كان العدد المقرر لبنود الاختبار هو مئة بند مثلاً كان من اللازم إعداد ما لا يقل عن 150 بند أو أكثر من ذلك لإفساح المجال أمام مصمّم الاختبار "لغربلتها" واختيار الأصلح منها في المراحل اللاحقة.

6) وضع تعليمات الاختبار بصورتها الأولى:

فالاختبار النفسي لابد أن يزوّد بتعليمات واضحة ووافية تحدّد مهمة المفحوص بدقة عند محاولته الإجابة عن أي من البنود التي يضمّها الاختبار وتمنع احتمال سوء فهم تلك المهمة. ولا شك أن المفحوص بحاجة إلى معرفة الوقت المعطى له للإجابة، وطريقة اختيار وتسجيل إجاباته عن البنود اللفظية، وكذلك طريقة التعامل مع البنود غير اللفظية. هذا بالإضافة إلى ما إذا كان سيسمح له بالتخمين أم لا عند الإجابة عن البنود التي تفسح المجال للتخمين

كبنود الاختيار المتعدّد مثلاً. وغني عن البيان أن تعليمات الاختبار هي متغير مهم له أثره في الأداء الاختباري سواء أكان بالاتجاه الإيجابي أم بالاتجاه السلبي وتعدّ التعليمات من المنبهات الأساسية للاختبار، ويؤدي غموضها إلى رفع مستوى القلق لدى المفحوصين وعدم تعاونهم في الموقف الاختباري. وتبعاً لذلك لابد أن تكون التعليمات واضحة للجميع، كما لابد من توحيدها في مواقف القياس الجمعية والفردية. وتخضع التعليمات كما تخضع البنود الاختبارية ذاتها للمراجعة والتعديل والتنقيح في ضوء نتائج التحليل الكيفي ونتائج التجارب الاستطلاعية والمبدئية التي يخضع لها الاختبار (التحليل الكمي).

7) وضع خطة تصحيح الاختبار:

من الأمور المهمة التي يتعين على مصمّم الاختبار أخذها بالحسبان طريقة التصحيح وما تتضمّنه من توزيع للدرجات على البنود الاختبارية المختلفة أو فئات منها انطلاقاً من الوزن النسبي لكل منها، وتقرير ما إذا كان سيستخدم معادلة التصحيح من أثر التخمين في البنود الموضوعية. هذا بالإضافة إلى تسجيل الإجابات وما إذا كان سيتمّ في كراسة الاختبار ذاتها أم في أوراق إجابة منفصلة، ثم ما إذا كان الاختبار سيصحح بالطريقة العادية أم بالآلة وما نوع المفتاح المستخدم في التصحيح.

وفيما يتصل بمسألة الحزر والتخمين يميل أكثر العاملين في القياس النفسي إلى ضرورة استخدام معادلة التصحيح من أثر التخمين توخّياً للدقة ولمنع المفحوص من التخمين الأعمى ومعاقبته إن لجأ إليه. وتظهر أهمية استخدام معادلة التصحيح من أثر التخمين حين تقل بدائل الإجابة (الاختيارات). فإذا كان المفحوص سيختار الإجابة الصحيحة من بين أربعة بدائل (اختيارات) كان احتمال نجاحه عن طريق التخمين يساوي 25٪، وإذا كانت بدائل الإجابة ثلاثة كان احتمال نجاحه عن طريق التخمين 33٪، وإذا كانت البدائل اثنين كان

احتمال نجاحه بالتخمين 50٪ وهكذا. هذا مع الإشارة إلى أن الباحثين قلما ينصحون باستخدام هذه المعادلة في الاختبارات الصفية وغير المقننة.

وتتلخّص الطريقة المتبعة في التصحيح من أثر التخمين في استخدام المعادلة التالية:

$$ع = ص - \frac{خ}{(ن-1)}$$

حيث يشير الرمز ع إلى العلامة المصححة.

ويشير الرمز ص إلى عدد البنود التي كانت الإجابة عنها صحيحة.

ويشير الرمز خ إلى عدد البنود التي كانت الإجابة عنها خاطئة.

ويشير الرمز ن إلى عدد البدائل في البند الواحد.

وإذا طبقت هذه المعادلة على بنود صواب – خطأ تكون العلامة المصححة هي عدد الإجابات الصحيحة مطروحاً منها عدد الإجابات الخاطئة.

وتفترض المعادلة السابقة أن العلامة التي حصل عليها المفحوص أعلى من علامته الحقيقية نظراً للجوئه إلى التخمين وتعاقبه تبعاً لذلك، ولا تعير أي اهتمام للبنود المتروكة (أي تلك البنود التي لم يجب عنها المفحوص التي سيخسر العلامة المخصصة لها بطبيعة الحال)، ذلك أن هذه البنود لا تدخل ضمن عدد الإجابات الخاطئة.

ومن الواضح أن الغرض من تطبيق المعادلة السابقة هو إقناع المفحوص بأن من الأفضل له ترك السؤال الذي يعجز عن الإجابة عنه بدلاً من الإجابة عنه بطريقة الحزر والتخمين الأعمى، لأنه في هذه الحالة الأخيرة سيصيب في بعض

الأسئلة ويخطئ في بعضها الآخر، ويمكننا عن طريق الإجابات الخاطئة تحديد ما ربحه بالتخمين ثم حذفه. وهذا يعني بعبارة أخرى أن المعادلة السابقة تعاقب المفحوص الذي يلجأ إلى التخمين بحسم عدد من الدرجات يقابل ما يحتمل أن يكسبه عن طريق التخمين.

وبين الانتقادات التي وجّهت للمعادلة السابقة هو أنها تنطلق من أن الإجابات الخاطئة جميعها تأتي بمحض التخمين، وهذا مجرد افتراض، حيث أن المفحوص قد يخطئ دون أن يخمّن. ثم إنَّ تطبيق هذه المعادلة يعقّد عملية التصحيح وقد يؤدي إلى ارتكاب بعض الأخطاء في التصحيح. هذا بالإضافة إلى أنها قد تثير تذمر واستياء المفحوصين. وقد أظهرت الدراسات أن استخدام هذه المعادلة لا يؤثّر، أو يؤثّر بدرجة محدودة، في تغيير مراكز الأفراد ضمن المجموعة.

ومن الباحثين من يقترح استخدام المعادلة التالية بديلاً عن المعادلة السابقة. وترتكز هذه المعادلة على تشجيع المفحوص على عدم التخمين (بترك البنود التي يتعذّر عليه الإجابة عنها) بدلاً من معاقبته إذا لجأ إلى التخمين، وذلك بإضافة ذلك الجزء من العلامة التي يتوقع أن يحصل عليها فيما لو لجأ إلى التخمين إلى عدد الإجابات الصحيحة. والمعادلة المقترحة هي:

$$ع = ص + \frac{م}{ن}$$

حيث يشير الرمز ع إلى العلامة المصححة.

ويشير الرمز ص إلى عدد البنود التي كانت الإجابة عنها صحيحة.

ويشير الرمز م إلى عدد البنود المتروكة (التي بقيت دون إجابة).

ويشير الرمز ن إلى عدد البدائل في البند الواحد.

والمعادلة الأخيرة تقوم على فلسفة مغايرة للمعادلة الأولى وهي مكافأة المفحوص الذي يحجم عن التخمين بمنحه عدداً من الدرجات يوازي ما يربحه الشخص الذي يلجأ إلى التخمين مما يلغي الفائدة التي قد يحققها هذا الأخير ويحقق التكافؤ بينهما من هذه الناحية. فإذا افترضنا أن أحد الطلاب أخضع لاختبار مؤلف من (100) سؤال من نوع صواب – خطأ واعتمد التخمين كلياً في الإجابة فإن علامته ستكون (50)، وأما زميله الذي امتنع كلياً عن الإجابة فإن علامته ستكون:

$$ع = ص + \frac{م}{ن}$$

حيث ع هي العلامة المصححة.

ص هي عدد الإجابات الصحيحة.

م هي عدد البنود المتروكة دون إجابة.

ن عدد البدائل في البند الواحد (الخيارات).

أي ع في هذه الحالة تساوي:

$$صفر + \frac{100}{2} = 50$$

أي أن الطالبين هنا يتعادلان.

غير أن تطبيق المعادلة السابقة سيرفع متوسط الدرجات وتباينها في المجموعة على الرغم من أنه لن يؤثر في ترتيب المفحوصين، أو انه سيؤثر في هذا الترتيب تأثيراً محدوداً. وبطبيعة الحال فإن ارتفاع متوسط الدرجات وتباينها لابد أن يرافقه ارتفاع في علامة الحد الأدنى الضروري للنجاح في الاختبارات

التحصيلية. فبدلاً من أن تكون هذه العلامة 50 من مئة مثلاً يجب رفعها إلى 70 أو أكثر. وهذا ما يتطلب استصدار قرارات إدارية خاصة بعلامة الحد الأدنى الضروري للنجاح في أسئلة الاختيار المتعدّد.

وثمة فريق آخر من الباحثين يقترح دمج المعادلتين السابقتين في معادلة واحدة بهدف معاقبة المفحوص إذا لجأ إلى التخمين، وتشجيعه على عدم اللجوء إليه من جهة ثانية.

والمعادلة الجديدة المقترحة هي:

$$ع = ص + \frac{خ}{(ن-1)} + \frac{م}{ن}$$

وتطابق دلالة الرموز في هذه المعادلة دلالتها في المعادلتين السابقتين.

ومن الجدير بالإشارة أن المعادلة الأولى من بين المعادلات الثلاث السابقة تصلح لامتحانات المرحلة الجامعية، كما تصلح بدرجة ما لامتحانات المرحلة الثانوية، في حين أن المعادلتين الثانية والثالثة تصلحان لمرحلة الدراسة الابتدائية بصورة خاصة.

8) التحليل الكيفي للبنود:

وهذا النوع من التحليل يرافق عملية إعداد البنود وتطويرها كما يكون خطوة لاحقة بها ومتمّمة لها. وهو يفيد في التأكد من أن البنود العديدة التي تمّ إعدادها ما هي إلا وسيلة لزيادة عدد المثيرات الدالّة على السمة أو القدرة المقيسة والمعبّرة عنها، كما يفيد في التأكد من أن هذه البنود العديدة والمتنوعة تمسّ أو تتناول الجوانب المختلفة للسمة أو القدرة المقيسة و" تعكسها " أو تمثّلها بحجمها الحقيقي. والواقع أن التحليل الكيفي للبنود والذي يتناول بخاصة شكل البنود، ومضمونها، وملاءمتها للأغراض الخاصة المرسومة، والتعليمات المرافقة لها يسهم

في تأسيس صدق المحتوى أو الصدق المنطقي للاختبار، ويضمن بصورة مسبقة التوصّل إلى هذا النوع من الصدق. ويتعذّر دون هذا التحليل العقلي المنطقي – والذي يفترض أن يبدأ مع عملية تصميم الاختبار ويواكبها – التوصل إلى مستوى عال من الصدق المنطقي، ذلك أن استخدام أي من الوسائل والأساليب التجريبية والإحصائية لا يجدي في التوصّل إلى هذا النوع من الصدق. هذا ولابد عند هذه الخطوة من خطوات بناء الاختبار من ترتيب البنود بصورة مبدئية حسب صعوبتها بحيث تبدأ بالأسهل وتنتهي بالأصعب، وذلك اعتماداً على خبرة المصمّم الخاصة وتقديره الشخصي لصعوبة البنود. وعموماً فإن التحليل الكيفي للبنود وترتيبها بصورة أولية يستهدفان تحضير الاختبار للتطبيق التجريبي. وبالانتهاء من هذه الخطوة يكون الاختبار بتعليماته وبنوده المختلفة قد أصبح جاهزاً للتطبيق التجريبي ثم للتحليل الكمي.

9) التحليل الكمي للبنود:

يتمّ هذا النوع من التحليل بعد إجراء التجربة الاستطلاعية والتجربة أو التجارب التمهيدية على عينات من الأفراد تشتق من المجتمع الأصلي للأفراد الذين صمّم الاختبار من أجلهم. ومن أغراض التجربة أو التجارب الاستطلاعية تعرّف الصعوبات الواقعية التي قد تعترض تطبيق الاختبار، وجمع الملاحظات عن البنود المختلفة، والتأكد من ملاءمتها للمفحوصين بالإضافة إلى التأكد من وضوح التعليمات، وملاءمة الوقت المخصص للاختبار ليصار في ضوء ذلك كله إلى إجراء بعض التعديلات الأولية اللازمة. وأما التجربة أو التجارب التمهيدية أو المبدئية فإنها تجري على عينات أوسع وأكثر تمثيلاً يصل عددها إلى مئات عدة من الأفراد أو يتجاوز ذلك في حالات، وتتركز أغراضها في الحصول على البيانات العددية اللازمة التحليل الكمي الإحصائي للبنود. واعتماداً على نتائج هذا التحليل يمكن تعرّف نقاط الضعف والقوة في الاختبار والكشف عن مستوى السهولة (والصعوبة) والتمييز لكل مفردة من مفرداته والتحقق نهائياً من وضوح

التعليمات وملاءمة البنود لما وضعت لقياسه. وفي ضوء ذلك كله تتمّ "غربلة" البنود أو تصفيتها واختيار أفضلها لإعداد الصورة النهائية للاختبار.

وتشمل عملية التحليل الكمي للبنود مجموعة من الإجراءات من أهمها تحديد مستوى الصعوبة (والسهولة) لكل بند من خلال استخراج معامل الصعوبة والسهولة الخاص بهذا البند بالإضافة إلى الكشف عن القدرة التمييزية للبنود (المفردات) أو ما يسمى بـ "صدق المفردات". ولاشك أن للتحليل الكمي لمفردات الاختبار أهميته القصوى. ففي ضوء نتائج هذا التحليل وبعد حساب معاملات السهولة (والصعوبة) والتمييز لكل مفردة يتمّ اختيار البنود التي ستؤلّف الاختبار بشكله النهائي كما أسلفنا. وبوسع الباحث أن يتحكّم إلى حد بعيد بمعامل سهولة الاختبار الكلي وبمتوسط درجاته وانحرافه المعياري وتباينه وإلى حد ما بثباته وصدقه باختيار البنود أو المفردات استناداً إلى معاملات السهولة (والصعوبة) والتمييز لكل منها كما سنرى.

استخراج معاملات السهولة والصعوبة:

يتمّ استخراج معامل سهولة البند عادة وفق الصيغة التالية:

$$\text{معامل سهولة البند أو السؤال} = \frac{\text{ص}}{\text{ص+خ}} \times 100$$

حيث يشير الرمز ص إلى عدد الإجابات الصحيحة، ويشير الرمز خ إلى عدد الإجابات الخاطئة.

وهذا يعني ببساطة أن معامل سهولة البند أو السؤال هو النسبة المئوية للإجابات الصحيحة عن هذا البند أو السؤال. مع الإشارة إلى أنه يمكن استخدام النسبة العشرية بدلاً من النسبة المئوية لاستخراج هذا المعامل. وفي هذه الحالة تستخدم الصيغة التالية:

$$\text{معامل سهولة البند أو السؤال} = \frac{\text{ص}}{\text{ص+خ}}$$

وأما العلاقة بين معامل السهولة ومعامل الصعوبة فهي علاقة عكسية ومباشرة، وللحصول على معامل الصعوبة يطرح معامل السهولة من واحد صحيح إذا استخدمت النسبة العشرية، ومن مئة إذا استخدمت النسبة المئوية. فإذا كان معامل سهولة السؤال هو 0.38 مثلاً (او 38٪) فإن معامل صعوبته هو: 1 – 0.38 = 0.62 أو 100٪ – 38٪ = 62٪. وهذا يعني ان 38٪ من المفحوصين الذين حاولوا الإجابة عن هذا السؤال أعطوا إجابات صحيحة عنه و 62٪ منهم كانت إجاباتهم خاطئة.

غير ان معاملات السهولة وكذلك معاملات الصعوبة لابد أن تتأثر بعامل التخمين في بنود الاختيار من متعدّد وبصورة خاصة حين يقل عدد بدائل الإجابة ويزداد احتمال الوصول إلى الجواب الصحيح عن طريق الحزر والتخمين. ولذلك ينصح الباحثون بتصحيح معاملات السهولة من أثر التخمين. ويمكن إجراء هذا التصحيح عن طريق دمج الصيغة الخاصة بحساب معامل السهولة مع الصيغة الخاصة بالتصحيح من أثر التخمين كما يلي:

$$\text{معامل السهولة} = \frac{\text{ص}}{\text{ص+خ}}$$

$$\text{والدرجة المصححة من أثر التخمين} = \text{ص} - \frac{\text{خ}}{\text{ن}-1}$$

$$\text{إذن معامل السهولة المصحح من أثر التخمين} = \frac{\text{ص} - \frac{\text{خ}}{\text{ن}-1}}{\text{ص+خ}}$$

فإذا كان عدد الإجابات الصحيحة عن السؤال 2، وعدد الإجابات الخاطئة 2، وعدد البدائل (الاختيارات) 4، فإن معامل سهولة هذا السؤال =

$$0.50 = \frac{2}{2+2}$$

وأما معامل السهولة المصحح من أثر التخمين لهذا السؤال فإنه:

$$\text{تقريباً } 0.33 = \frac{\frac{2}{3}-2}{4} = \frac{\frac{2}{1-4}-2}{2+2}$$

غير أن استخدام معاملات السهولة والصعوبة المحسوبة بطريقة النسب في ترتيب بنود الاختبار حسب صعوبتها لا يراعي التوزيع الاعتدالي "المعياري" للقدرات، ولا يخرج في الواقع عن المستوى الترتيبي أو الرتبي للقياس (المقياس الرتبي) والذي يعطي ترتيباً " فظاً "للبنود يدلّ على الاختلاف بينها من حيث الصعوبة دونما تحديد لمقدار هذا الفرق أو الاختلاف بصورة منتظمة ودقيقة. وهذا يعني بعبارة أخرى أن الفروق بين الرتب التي تأخذها تلك البنود غير متساوية، فالفرق في الصعوبة بين 25٪ و35٪ لا يساوي الفرق في الصعوبة بين 60٪ و 70٪ مثلاً مع انه يبلغ 10٪ في الحالتين. وتبعاً لذلك فإن معاملات السهولة المحسوبة بطريقة النسب لا تصلح في الواقع إلا لترتيب المفردات بصورة أولية وذلك " لعجزها عن تحديد الفروق القائمة بين مراتب سهولة تلك المفردات مع أن لهذه الفروق أهميتها في الاختيار النهائي للمفردات وفي التدريج المنتظم للسهولة " (البهي السيد ،1978، ص450). لهذا عمد الباحثون إلى تحويل تلك المعاملات أو النسب إلى درجات معيارية بهدف الوصول إلى مقياس أفضل لمستوى الصعوبة وذلك من منطلق أن كل نسبة من تلك النسب تدل على نسبة مساحة معينة إلى المساحة الكلية لمنحني التوزيع الاعتدالي، وبالتالي يمكن تحويلها إلى الدرجة المعيارية المقابلة لها في هذا المنحني. وبالرجوع إلى جداول

مساحات المنحني الاعتدالي المعياري يمكن الحصول على الدرجات المعيارية المقابلة للنسب المختلفة، وبذلك يتمّ الحصول على المعاملات المعيارية للسهولة.

غير أن الدرجات المعيارية تعاني من كثرة الإشارات السالبة. لذا عمد الباحثون إلى تعديل تلك الدرجات والحصول بالتالي على المعاملات المعيارية المعدلة للسهولة. ومن التعديلات المقترحة في هذا المجال إضافة خمس درجات معيارية موجبة إلى العدد الكلي للدرجات المعيارية، مما يتيح قلب الدرجات المعيارية السالبة وهي خمس درجات في التوزيع السوي إلى درجات موجبة والتخلص كلياً من الإشارات السالبة. وبدلاً من أن يتراوح التوزيع من – 5 إلى + 5 درجات معيارية مروراً بالصفر يتراوح التوزيع الجديد من صفر إلى +10 ويصبح متوسطه 5 بدلاً من الصفر. وعليه فإذا كان معامل السهولة المعياري هو – 0.91 مثلاً أي تحت المتوسط بأقل من انحراف معياري واحد يكون معامل السهولة المعياري المعدل وفقاً للتوزيع الجديد هو:

$$4.11 = 5 + 0.91 -$$

وأما معامل الصعوبة المعياري المعدل لهذا البند فهو:

$$5.91 = 4.11 - 10$$

ومن التعديلات التي وضعتها هيئة خدمات الاختبارات التربوية الأمريكية استخدام توزيع آخر جديد للدرجات المعيارية يتراوح مداه من 1 إلى 25 ويكون متوسطه 13 وذلك باستخدام المعادلة التالية:

$$\text{د م} = -4\text{ع} + 13$$

حيث يدل الرمز د م على معامل الصعوبة المعياري المعدل.

ويرمز الرمز 4 ع وهو عدد ثابت إلى عدد وحدات الانحراف المعياري وهي أربع وحدات في هذه الحالة بدلاً من خمس وحدات في الحالة السابقة.

فإذا كان البند متوسط الصعوبة وأجاب عنه 50 ٪ من الأفراد أي يقع تماماً في وسط المنحنى الاعتدالي فإن معامل الصعوبة المعياري المعدل لهذا البند هو:

$$13 = 13 + \text{صفر} \times 4 -$$

وإذا كان البند شديد السهولة وأجاب عنه 99.8٪ من الأفراد، وهذه النسبة تقع على مسافة 3 انحرافات معيارية عن المتوسط (+3 ع) فإن معامل الصعوبة المعياري المعدل لهذا البند هو:

$$1 = 13 + 4 - \times 3 +$$

ونحصل على نتيجة معاكسة لذلك تماماً إذا كان البند شديد الصعوبة وأجاب عنه فقط 1٪ من المفحوصين وهي النسبة التي تقع على مسافة 3 انحرافات معيارية عن المتوسط (− 3 ع) ويكون معامل الصعوبة المعياري المعدل في هذا الحالة هو:

$$25 = 13 + 4 - \times 3 -$$

هذا ويمكن استخدام الطرائق السابقة المتبعة في حساب معاملات سهولة (وصعوبة) المفردات التي يضمّها الاختبار في حساب معامل سهولة (وصعوبة) الاختبار ككل مباشرة ودون حساب معامل سهولة وصعوبة كل مفردة من مفردات الاختبار على حدة. فكما أن معامل سهولة البند يحسب بقسمة مجموع الدرجات المتحصّلة على الإجابات الصحيحة عن هذا البند على المجموع الكلي لدرجات هذا البند، فإن معامل سهولة الاختبار الكلي يمكن حسابه بقسمة مجموع

الدرجات المتحصّلة على الإجابات الصحيحة عن الاختبار بأكمله على المجموع الكلي لدرجات هذا الاختبار. غير أن معامل سهولة (وصعوبة) الاختبار ككل هو مؤشر عام لمستوى صعوبة الاختبار ولا يكشف عن مدى هذه الصعوبة وانتشارها في البنود المختلفة، ولا يفيد بالتالي في عملية "تنقية" الاختبار واختيار البنود استناداً إلى مستوى صعوبتها واستبعاد البنود شديدة السهولة أو شديدة الصعوبة.

ويحسب معامل سهولة الاختبار ككل بالصيغة التالية:

$$\text{معامل سهولة الاختبار} = \frac{\text{مجموع درجات المفحوصين}}{\text{النهاية العظمى للدرجات} \times \text{عدد المفحوصين}}$$

فإذا كان عدد المفحوصين 10

وكانت النهاية العظمى للدرجات 100

وكان مجموع الدرجات التي حصل عليها المفحوصون 700

فإن معامل سهولة هذا الاختبار هو:

$$0.70 = \frac{700}{10 \times 100}$$

وبالرجوع إلى جداول مساحات المنحنى الاعتدالي المعياري يمكن الحصول على الدرجة المعيارية المقابلة لمعامل سهولة قدره 0.70، وهي معامل السهولة المعياري والذي يبلغ في هذه الحالة +0.524 (أي يزيد بأكثر من نصف درجة على منتصف المنحنى الاعتدالي).

كما يمكن تعديل هذا المعامل تماماً كمعامل سهولة المفردات وذلك بإضافة +5 درجات معيارية فيصبح 5.524.

فوائد معاملات السهولة والصعوبة:

يمكن إجمال فوائد استخراج معاملات السهولة والصعوبة فيما يلي:

1. **ترتيب بنود الاختبار:**

فاستناداً إلى تلك المعاملات يمكن ترتيب البنود التي سيضمّها الاختبار بحيث تبدأ بالأسهل وتنتهي بالأصعب، ولهذا الترتيب أهميته من حيث انه يتيح استثارة دافعية المفحوص وينمّي ثقته بنفسه بإجابته عن البنود السهلة أولاً، ويضعف احتمال تعرض المفحوص "للصدمة" منذ بداية عمله في الاختبار، كما يمنع احتمال إضاعة الكثير من الوقت والجهد في بنود قد يعجز المفحوص عن الإجابة عنها أصلاً لأنها تتجاوز حدود أو سقف قدرته. غير أن ترتيب البنود حسب مبدأ التدرج في الصعوبة لا يعني القبول بمدى واسع من الصعوبة. ويرى علماء القياس أن مدى الصعوبة بين أسهل البنود وأصعبها يجب أن يكون محدوداً بقدر الإمكان في الاختبارات المقننة ويحيث تدور أكثر معاملات السهولة حول نسبة 50٪ والقليل منها فقط يقع في المدى من 40٪ إلى 60 ٪ او يتجاوز هذا المدى بقليل.

2. **انتقاء بنود الاختبار:**

فانتقاء البنود لضمها إلى الصورة النهائية للاختبار أو ما يعرف بعملية "تنقية " الاختبار يستند أيضاً إلى مستوى الصعوبة. وفي هذا الصدد يشير علماء القياس إلى أن المستوى الأكثر ملاءمة لصعوبة البنود هو 50٪ كما ذكرنا. فالبنود التي يصل معامل سهولتها أو صعوبتها إلى 0.5 او 50٪ تنتج اكبر قدر ممكن من التباين أو التمايز بين الأفراد وتكون حساسة للفروق الدقيقة بينهم إلى الدرجة القصوى. ومن المعلوم أن التباين ما هو إلا حاصل ضرب معامل السهولة بمعامل الصعوبة. ويمكن الوصول إلى الحد الأقصى للتباين حين يكون معامل سهولة البند 0.50 وذلك على النحو التالي:

إذا كان التباين = معامل السهولة × معامل الصعوبة

فإن الحد الأقصى للتباين = 0.50 × 0.50 = 0.25

وإذا زاد معامل السهولة أو نقص عن 0.50 تعذّر الحصول على الحد الأقصى للتباين. فإذا كان معامل سهولة البند 0.6 مثلاً كانت القيمة العددية للتباين هي:

0.6 × 0.4 = 0.24 وهذا التباين أقل من 0.25

وإذا كان معامل سهولة البند 0.8مثلاً كانت القيمة العددية للتباين هي:

0.8 × 0.2 = 0.16 وهي أقل أيضاً من 0.25

وإذا كان معامل سهولة البند 0.9مثلاً كانت القيمة العددية للتباين هي:

0.9 × 0.1 = 0.09 وهي أقل أيضاً من 0.25

ويتّضح من الأمثلة السابقة أن القيمة العددية للتباين تتناقص كلما ابتعدنا صعوداً أو هبوطاً عن 0.5. غير أنه يستحيل توحيد صعوبة البنود جميعها عند مستوى 0.5 تماماً. ولذا ينصح الباحثون بالإكثار من الأسئلة المتوسطة في سهولتها وصعوبتها، والتقليل من الأسئلة السهلة والصعبة، والاستغناء كلياً عن الأسئلة شديدة السهولة أو الصعوبة إذا كان الغرض من الاختبار هو الكشف عن الفروق الدقيقة بين الأفراد وإظهار القدر الأعلى من التباين أو التمايز بينهم. وهذا الأمر يتعارض مع الفكرة الشائعة والقائلة:

إن الاختبار الأقدر على التمييز بين الأفراد هو ذلك الاختبار الذي يتضمّن "تشكيلة" واسعة من الأسئلة بينها الأسئلة الصعبة التي تتحدى الأقوياء والأسئلة السهلة الموجهة للضعفاء بالإضافة إلى الأسئلة المتوسطة.

3. بناء الشكل المكافئ أو عدد من الاختبارات المتكافئة:

تفيد معاملات السهولة والصعوبة في بناء الشكل المكافئ للاختبار أو عدد من الاختبارات المتكافئة. والواقع أن التكافؤ أو التعادل في مستوى سهولة البنود أو صعوبتها وفي مستوى سهولة الاختبار او صعوبته ككل يعدّ الشرط الأساسي في الحصول على اختبارات متكافئة. وحين يعمد الباحث إلى تطوير اختبارات متكافئة يتمّ اختيار البنود بحيث يكون معامل سهولة البند الأول في الاختبار الأول مساوياً أو قريباً من معامل سهولة البند الأول في الاختبار الثاني الذي اختير ليكون مناظراً له من حيث الشكل والمحتوى، كما يكون معامل سهولة البند الثاني في الاختبار الأول مساوياً أو قريباً من معامل سهولة البند الثاني في الاختبار الثاني الذي اختير بدوره ليكون مناظراً له من حيث الشكل والمحتوى وهكذا .... وللاختبارات المتكافئة أهميتها في حساب ثبات الاختبار، كما تتيح الفرصة لاستخدام أكثر من شكل واحد للاختبار، وتظهر فائدتها بخاصة حين تواجَه عملية تطبيق الاختبار "الأصلي" بعراقيل معينة أو يشكّك بنتيجة المفحوص في ذلك الاختبار.

4. حساب الانحراف المعياري والتباين مباشرة من معاملات السهولة والصعوبة:

والصيغة الخاصة بحساب الانحراف المعياري للبند هي:

$$\text{الانحراف المعياري للبند} = \sqrt{\text{معامل السهولة} \times \text{معامل الصعوبة}}$$

وأما الصيغة الخاصة بحساب التباين (وهو مربع الانحراف المعياري) فهي:

$$\text{التباين} = \text{معامل السهولة} \times \text{معامل الصعوبة}$$

فإذا كان معامل سهولة البند أو السؤال هو 0.7 فإن قيمة الانحراف المعياري لهذا السؤال هي:

$$0.458 = \sqrt{0.21} = \sqrt{0.3 \times 0.7}$$

وأما قيمة التباين لهذا السؤال فهي:

$$0.21 = 0.3 \times 0.7$$

ويوسع القارئ استخدام الصيغة العامة لحساب الانحراف المعياري وكذلك التباين لمجموعة من الدرجات واستخدام الصيغة القائمة على معامل السهولة، وسيجد أن النتيجة التي سيحصل عليها في الحالتين واحدة. ويظهر ذلك واضحاً في الجدول التالي:

الجدول رقم (1): حساب الانحراف المعياري لدرجات أحد الأسئلة:

الأفراد	الدرجات	مربعات الدرجات
1	1	1
2	1	1
3	صفر	صفر
4	1	1
5	1	1
6	صفر	صفر
7	صفر	صفر
8	1	1
9	صفر	صفر
10	1	1
10	مجموع الدرجات = 6	مجموع مربعات الدرجات = 6
	المتوسط = $\frac{6}{10}$ = 0.6	متوسط مربعات الدرجات = 0.6

وبتطبيق المعادلة العامة لحساب الانحراف المعياري وهي:

الانحراف المعياري = $\sqrt{\text{متوسط مربعات الدرجات} - \text{مربع متوسط الدرجات}}$

يكون الانحراف المعياري لهذا السؤال هو:

$\sqrt{0.6 - (0.6)^2} = 0.49$ تقريباً

وبتطبيق المعادلة القائمة على معامل السهولة:

يكون الانحراف المعياري لهذا السؤال (والذي يبلغ معامل سهولته 0.6) هو:

الانحراف المعياري للبند = $\sqrt{0.6 \times 0.4} = \sqrt{0.24} = 0.49$ تقريباً

وهي النتيجة السابقة نفسها.

ولابد من الإشارة إلى أن معاملات السهولة والصعوبة ليست أرقاماً نهائية أو ثابتة فهي تتحدّد بأداء أفراد العينة التجريبية وبجملة الشروط الواقعية التي يطبق ضمنها الاختبار، ومن الطبيعي أن تكون حساسة لأي تغيير يطرأ على العينة أو على أي من الشروط المحيطة بعملية تطبيق الاختبار. ولتلك المعاملات أهميتها في اختبارات القدرات والتحصيل المقننة على وجه التحديد، ويتمّ في ضوئها تعديل صعوبة الاختبار حتى يعطي في النهاية توزيعاً أقرب ما يكون إلى التوزيع السوي أو الاعتدالي. لذلك كثيراً ما تحذف بنود معينة أو تعدّل وتضاف بنود صعبة أو سهلة، كما تعاد هذه العملية لأكثر من مرة، حتى يتمّ الوصول إلى التوزيع السوي أو أقرب ما يكون إليه. بالإضافة إلى ذلك فإن لاختيار البنود استناداً إلى معاملات السهولة والصعوبة أثره المباشر في رفع مستوى الصدق والثبات كما سنرى.

صدق المفردات (استخراج معاملات التمييز):

لا تقتصر عملية التحليل الكمي الإحصائي للبنود على استخراج معاملات السهولة والصعوبة بل تشمل بالإضافة إلى ذلك استخراج معاملات التمييز لتلك البنود وهو ما يدخل في نطاق ما يسمى "صدق المفردات". والواقع أن مستوى صعوبة البنود يؤثّر تأثيراً مباشراً في قدرتها التمييزية كما أشرنا، وقد يكون من الأنسب، تبعاً لذلك، عدم الفصل بينهما بل والنظر إليهما على أنهما جانبان لعملية واحدة متكاملة تستهدف اختيار أفضل البنود وأكثرها ملاءمة استناداً إلى مؤشرات الصعوبة والتمييز معاً. والمقصود بالقدرة التمييزية للبنود حساسيتها للفروق الدقيقة بين الأفراد في السمة أو القدرة المقيسة، وأما معامل التمييز فهو مفهوم كمي إحصائي يعبر بلغة العدد عن درجة تلك الحساسية ومدى قدرة البند على التمييز أو التفريق بين الأفراد في ذلك الجانب أو المظهر من السمة الذي يتصدّى لقياسه.

ولاشك في أن القدرة التمييزية للبنود تتصل مباشرة بصدق تلك البنود أو المفردات ونجاحها في قياس ما وضعت لقياسه، وتعبر عن درجة هذا الصدق. ومن الباحثين من يوحّد بين القدرة التمييزية والصدق وينظر إليهما على انهما مترادفان أو اسمان لمسمى واحد. وعلى أية حال فإن صدق الاختبار بأكمله يعتمد مباشرة على القدرة التمييزية لبنوده. فإذا ما سعى الباحث إلى تحقيق قدر عال من الصدق لاختباره ككل تعيّن عليه دراسة القدرة التمييزية لكل من البنود التي يضمّها على حدة. هذا مع الإشارة إلى أنه قد يكون من الأنسب النظر إلى دراسة القدرة التمييزية واستخراج معاملات التمييز على أنها مظهر من مظاهر التحقق من صدق المفردات ولا "تستقطب" هذا النوع من الصدق برمته نظراً لوجود أساليب أخرى للتحقق من هذا الصدق. ومن هذه الأساليب دراسة الارتباط بين البند والمحكّ (أو الميزان) الخارجي أو دراسة الارتباط بين البند وبقية الاختبار وهو محك أو ميزان داخلي وذلك لبيان مدى قدرة البند على التنبؤ بالمحكّ الخارجي في الحالة الأولى، وبيان ما إذا كان البند بوصفه مفردة أو وحدة من

وحدات الاختبار يقيس ما يقيسه الاختبار ككل ويعمل بالاتجاه نفسه الذي يعمل به الاختبار بأكمله، أو يتّسق معه في الحالة الثانية.

وثمة طرائق عديدة لحساب صدق المفردات واستخراج مؤشر القدرة التمييزية لا يتسع المقام للوقوف عندها جميعاً، ونكتفي باستعراض الطرائق التالية الأكثر أهمية وشيوعاً بينها:

1) حساب صدق المفردات باستخدام معاملات الارتباط:

وتقوم هذه الطريقة على دراسة الارتباط بين أداء المفحوصين على الاختبار ككل وادائهم على كل بند من بنوده على حدة لبيان مدى "تمشي" البند مع الاختبار ككل واتساقه معه. ومن الصيغ المستخدمة لحساب الصدق بهذه الطريقة الصيغة التالية:

$$\text{ر ث} = \frac{\text{م أ} - \text{م ب}}{\text{ع}} \sqrt{\text{أ} \times \text{ب}}$$

حيث يشير الرمز رث إلى معامل الارتباط الثنائي.

والرمز م أ إلى متوسط الصواب.

والرمز م ب إلى متوسط الخطأ.

والرمز أ إلى نسبة الصواب.

والرمز ب إلى نسبة الخطأ.

والرمز ع إلى الانحراف المعياري لدرجات المحكّ أو الميزان.

ويفيد تنظيم الجدول التالي في حساب معامل الارتباط بين الاختبار الكلي وأحد بنوده بحسب الصيغة السابقة. ويدل العمود الأول في هذا الجدول على

درجات المفحوصين في الاختبار الكلي بعد أن رتّبت ترتيباً تصاعدياً، ويدل العمود الأخير على تكرار درجات المفحوصين في الاختبار الكلي. وأما العمود الثاني فيدل على تكرار الإجابات الصحيحة عن السؤال في حين أن العمود الثالث يدل على تكرار الإجابات الخاطئة عن السؤال. وذلك على النحو التالي:

الجدول رقم (2): حساب معامل الارتباط الثنائي بين البند والاختبار الكلي:

درجات المفحوصين في الاختبار الكلي	تكرار صواب السؤال الأول	تكرار خطأ السؤال الأول	تكرار درجات المفحوصين في الاختبار الكلي
11	0	1	1
12	0	2	2
13	2	1	3
14	0	1	1
15	2	1	3
16	3	1	4
17	2	1	3
18	1	0	1
19	1	0	1
20	1	0	1
	عدد الإجابات الصحيحة =12 مجموع الدرجات=195	عدد الإجابات الخاطئة =8 مجموع الدرجات = 110	عدد المفحوصين =20 مجموع الدرجات =305
	المتوسط = $\frac{195}{12}$	المتوسط = $\frac{110}{8}$	المتوسط= $\frac{305}{20}$
	16.25=	13.75 =	15.25 =
	النسبة = $\frac{12}{20}$ 0.6=	النسبة = $\frac{8}{20}$ 0.4=	الانحراف المعياري 2.36 =

ويتبين من هذا الجدول أن

م ا = 16.25 ، ا = 0.6

م ب = 13.75 ، ب = 0.4

ع = 2.36

إذن ر = $\frac{16.25-13.75}{2.36} \times \sqrt{0.6 \times 0.4}$

$= \frac{2.5}{2.36} \times \sqrt{0.24}$

$= 1.059 \times 0.489 = 0.52$ تقريباً

ويرى الباحثون أن هذه الطريقة تعدّ من أدق الطرائق لحساب صدق المفردات، غير انها تتطلب جهداً كبيراً ووقتاً طويلاً ولاسيما إذا كثر عدد بنود الاختبار.

ومن الباحثين من اقترح حساب صدق المفردات عن طريق تقسيم المفحوصين إلى فئتين أو أكثر، تضمّ الأولى منهما أولئك الأفراد الذين حصلوا على أعلى الدرجات، وتضمّ الثانية أولئك الأفراد الذين حصلوا على أدنى الدرجات في الاختبار ككل، ثم مقارنة درجات السؤال في الفئة العليا بدرجات هذا السؤال نفسه في الفئة الدنيا، فإذا أظهرت المقارنة أن درجات السؤال في الفئة العليا أعلى منها في الفئة الدنيا دلّ ذلك على أن السؤال يميّز بينهما، ويماشي الاختبار ككل في قياس ما وضع له. وكلما كان الفارق كبيراً بين الفئتين ارتفع الصدق بطبيعة الحال، وكلما نقص هذا الفارق انخفض الصدق. فإذا انعدم دلّ ذلك على أن السؤال لا يميّز بين الفئتين.

وقد لجأ بعض الباحثين إلى التقسيم الثلاثي للمفحوصين وإجراء المقارنة بين الفئتين العليا والدنيا فقط (الثلث الأعلى والثلث الأدنى) لدورهما القوي في المقارنة، واستبعاد درجات أفراد الفئة الوسطى نظراً لدورها الضئيل والمحدود في المقارنة. واقترح كيللي أن يقتصر عدد أفراد كل من الفئتين العليا والدنيا على نسبة 27٪ فقط من افراد العينة واستبعاد درجات أفراد الفئة الوسطى المتبقية التي تشكل 46٪ للسبب ذاته. ومن الباحثين من ارتأى إمكان استخدام أي نسبة تتراوح من 25٪ إلى 33٪ لتؤلّف كلاً من الفئتين العليا والدنيا.

وبالاستعانة بمعاملات السهولة لدى كل من الفئتين العليا والدنيا قام فلاناجان بحساب معاملات ارتباط الاختبار بكل من المفردات التي يضمّها.

يشير معامل السهولة العلوي إلى نسبة الإجابات الصحيحة عن البند لدى أفراد الفئة العليا فقط (ونسبتها إلى العدد الكلي لأفراد العينة 27٪).

وأما معامل السهولة السفلي فيشير بالمقابل إلى نسبة الإجابات الصحيحة عن البند لدى أفراد الفئة الدنيا أو السفلى فقط (ونسبتها إلى العدد الكلي لأفراد العينة هي أيضاً 27٪). وبحساب معاملات السهولة العلوية والسفلية لكل مفردة والرجوع إلى الجداول التي أعدها فلاناجان بهذا الخصوص يمكن معرفة معامل ارتباط هذه المفردة بالاختبار ككل. فإذا كان عدد الذين أعطوا إجابات صحيحة عن أحد البنود في الفئة العليا هو 22 وفي الفئة الدنيا هو 9 فإن معامل السهولة العلوي لهذا البند هو:

$$0.81 = \frac{22}{7}$$

وأما معامل السهولة السفلي لهذا البند فهو:

$\frac{9}{27}$ = 0.33 تقريباً. ويقابل هذان المعاملان للسهولة في جداول فلاناجان.

معامل ارتباط قدره 0.49 (انظر: الجداول الإحصائية للبهي السيد، ص70)

2) حساب صدق المفردات بطريقة الفروق الطرفية:

اقترح جونسون صيغة أخرى لحساب صدق المفردات تعتمد بدورها على معاملات السهولة في الفئتين العليا والسفلى ولكن تستغني عن حساب معاملات الارتباط ولا تتطلب سوى عملية طرح بسيطة لمعامل السهولة السفلي من معامل السهولة العلوي. وتظهر هذه الصيغة في المعادلة التالية:

معامل صدق السؤال (معامل التمييز) = $\frac{ص_ع - ص_س}{0.27\ ن}$

حيث يدل الرمز $ص_ع$ على عدد الإجابات الصحيحة في الفئة العليا.

ويدل الرمز $ص_س$ على عدد الإجابات الصحيحة في الفئة السفلى.

ويدل الرمز ن على عدد المفحوصين.

ويمكن أن تأخذ المعادلة السابقة الشكل البسيط التالي:

معامل صدق السؤال = معامل السهولة العلوي – معامل السهولة السفلي

فإذا عدنا إلى المثال السابق واستخدمنا هذه الصيغة السهلة والبسيطة في حساب معامل الصدق فإن قيمة هذا المعامل ستكون:

$$0.48 = 0.33 - 0.81$$

وهـذه القيمـة قريبـة جـداً مـن قيمـة معامـل الصـدق المحسـوب بطريقـة فلاناجـان والبالغـة 0.49. ولـو عـدنا أيضـاً إلى المثـال المحسـوب بطريقـة معامـل الارتباط الثنائي (الجدول رقم (2)) واستخدمنا هذه الصيغة الجديدة في حساب معامل الصدق لحصلنا على قيمة لهذا المعامل قريبة من القيمة التي حصلنا عليها بطريقة الارتباط الثنائي.

وينصح الكثيرون باستخدام هذه الطريقة لبساطتها وسهولة حسابها.

هـذا ويـتراوح معامـل صـدق السـؤال مـن +1 إلى – 1 مـروراً بالصـفر الـذي يدل على انعدام القدرة التمييزية للسؤال، وبصورة عامة كلما ارتفع هذا المعامل كـان السـؤال أفضـل، والسـؤال الـذي يقـترب معامـل تمييـزه مـن 0.40 أو يتجـاوز ذلك يعدّ مناسباً، وأما السؤال الذي يقل معامل تمييزه عن ذلك بشكل ملحوظ فقد يحتاج إلى تعديل أو حذف. من جهة أخرى فقد أشار جونسون إلى إمكان استخدام معامـل سـهولة السـؤال لـدى كـل مـن الفئـتين العليـا والـدنيا فقـط في حساب معامل سهولة السؤال مباشرة على نطاق المجموع الكلي لأفراد العينة وإهمال درجات الفئة الوسطى، وذلك باستخدام المعادلة البسيطة التالية:

$$\text{معامل سهولة السؤال} = \frac{\text{معامل السهولة العلوي} + \text{معامل السهولة السفلي}}{2}$$

فإذا كان معامل سهولة السؤال لدى الفئة العليا 0.80

وكان معامل سهولة السؤال لدى الفئة الدنيا 0.30

فإن معامل سهولته هو: $\frac{0.30+0.80}{2} = 0.55$

ويـوفّر حسـاب معامـل سـهولة السـؤال بهـذه الطريقـة الكـثير مـن الوقـت والجهد على حساب القليل من الدقة كما يرى جونسون. وعموماً فإن معاملات السهولة والتمييز ليست كماً ثابتاً مطلقاً، وبدلاً من أن نقول: إن معامل سهولة

احد الأسئلة بلغ 0.58 ومعامل تمييزه بلغ 0.43 مثلاً من الأفضل ان نقول: إن معامل سهولة هذا السؤال بلغ 0.58 ومعامل تمييزه بلغ 0.43 عندما طبق الاختبار في ظرف معين على عينة من الأفراد ذات مواصفات محددة.

وعموماً يتضمّن صدق المفردات إجراء دراسة مفصلة لبنود أو مفردات الاختبار من البداية إلى النهاية وذلك من خلال دراسة الإجابات الصحيحة والخاطئة عن كل منها وقدرتها التمييزية ثم بيان ما إذا كانت تعمل أو تسير بالاتجاه نفسه الذي يسير به الاختبار بأكمله. ومن هذه الزاوية فإن تحليل البنود على النحو السابق يقدّم دليلاً حول الاتساق الداخلي للاختبار أو ما يسمى "الصدق الداخلي " ويظهر أن البنود مجتمعةً "متجانسة" إلى حد بعيد وتقيس شيئاً ما واحداً ولا يدل على أنها تقيس ذلك الشيء الذي وضعت لقياسه بالضرورة. وهذا يعني أن البيانات المتحصّلة من تحليل البنود لا يمكن عدّها بحد ذاتها دليلاً كافياً على الصدق. " وهذه البيانات يمكن أن تدعم الصدق وتؤكده حين يتمّ إثبات صدق الاختبار الكلي بطريقة ما أو يفترض هذا الصدق منطقياً " كما يقول جرونلند(Gronlund,1971,p.11). وعلى ذلك فإن السؤال الأهم حول فاعلية البنود لا يرتبط بمستوى صعوبتها وقدرتها التمييزية فقط بل يرتبط أيضاً بقياسها لتلك السمة أو لذلك الجانب من السمة الذي وضعت لقياسه أيضاً. ومن الطبيعي إذا كان الاختبار يتصدّى لجوانب متنوعة من السمة المقيسة، كأن يتصدّى للجانبين اللفظي والأدائي للذكاء معاً، أن يشتمل على طوائف متنوعة من البنود تتصدّى كل منها لأحد جوانب السمة المقيسة. ويتوقع أن تترابط هذه الطوائف مع بعضها ترابطاً موجباً بطبيعة الحال دون أن يؤدي هذا الترابط إلى إلغاء استقلاليتها النسبية في إطار ما وضعت لقياسه.

والواقع أن تحليل البنود – وهو إجراء تجريبي إحصائي بعدي يتمّ بعد إعداد الاختبار بصورته الأولية وتطبيقه على عينات من المفحوصين – لا يغني عن صدق المحتوى وهو صدق منطقي قبلي يسبق تطبيق الاختبار، أو الصدق البنيوي (الافتراضي) الذي يجمع بين الأساليب التجريبية وغير التجريبية في تحقيق

الصدق. والإفراط أو المبالغة في عملية التحليل الإحصائي لبنود الاختبار قد لا تكون مجدية وفعالة في كل الأحيان ولاسيما إذا كانت البنود الموضوعة ضعيفة أساساً وتفتقر إلى الصدق المنطقي. وهذا يؤكد ما أشرنا إليه سابقاً حول دور شخصية مصمّم الاختبار و "براعته" الخاصة في تطوير بنود ذات مواصفات عالية منذ البداية، وقبل أن يوضع الاختبار على "محك" التجريب ويخضع للتحليل الكمي الإحصائي.

تقنين الاختبار:

يشير مصطلح التقنين إلى توحيد الشروط المحيطة بالمفحوصين كافة و"ضبط" العوامل والمتغيرات جميعها التي يمكن أن تؤثّر في أدائهم الاختباري وتحييد أثرها إلى الدرجة القصوى، وذلك بهدف قياس المتغير الذي صمّم الاختبار لقياسه بحجمه الحقيقي لدى تلك العينة من الأفراد التي سيطبق عليها ودون أن يكون لأي من المتغيرات الأخرى أثره في أداء أفراد تلك العينة. وقد يكون من الخطأ النظر إلى التقنين على أنه خطوة أخيرة ومتممة لعملية بناء الاختبار أو لاحقة بها، والأصح أن تنظر إليه على أنه عملية مواكبة لها أو جزء لا يتجزأ منها. والواقع أن التقنين بالمعنى الشامل للكلمة يعتمد على رسم خطة شاملة وواضحة ومحددة لجميع خطوات تصميم الاختبار وإجراءاته وطريقة تطبيقه وتصحيحه وتفسير درجاته، وتحديد السلوك المطلوب من المفحوص والشروط المحيطة به في أثناء تطبيق الاختبار، بالإضافة إلى وجود معايير لتفسير الدرجات.

ويميّز علماء القياس بين معنيين للتقنين وهما:

أولاً: أن تكون تعليمات الاختبار وصياغة بنوده والزمن المخصص له وشروط تطبيقه وطريقة تصحيحه موحدة في كل المواقف وبما يسمح بإمكان الحصول على النتائج نفسها في حال إعادة إجراء الاختبار. ومن هذه الناحية يعني

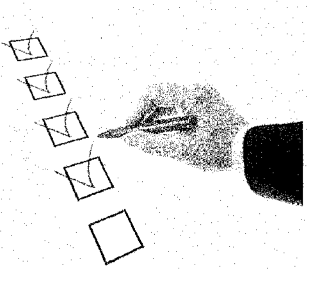
التقنين التوحيد. ويفقد الاختبار أساسه العلمي والموضوعي إذا لم يكن مقنناً بهذا المعنى.

ثانياً: أن يخضع الاختبار للتقنين من خلال تطبيقه على عينة ممثلة للمجتمع الأصلي بهدف استخراج معايير معينة تحدّد معنى الدرجة الخام التي يحصل عليها الفرد في ضوء تمركز وتشتّت درجات أفراد عينة التقنين (أو التعيير). وغني عن البيان أنه لا يمكن أن يكون الاختبار مقنناً بالمعنى الثاني دون أن يكون مقنناً بالمعنى الأول، وأنّ عملية استخراج معايير الاختبار، وهي المرحلة الأخيرة من مراحل تطويره، تحقق فوائد وأغراضاً عديدة منها تعرّف الوضع النسبي للفرد ضمن المجموعة، وتفسير الدرجة التي يحصل عليها، بالإضافة إلى ذلك فإنّ هذه العملية – عملية استخراج المعايير – هي عملية كبيرة ومجهدة من الناحية التجريبية والإحصائية وتستلزم تعاون عدد غير قليل من المختصين والخبراء، كما أنها مكلّفة مادياً وتستغرق وقتاً غير قليل.

ومن المفيد الإشارة هنا إلى أن الاختبار المقنن بالمعنى الأول، أي الذي أخضع للتقنين دون أن تستخرج معاييره، هو أداة قياس علمية ودقيقة ويمكن أن تحقق أغراضاً وفوائد عديدة في مجالات التشخيص والبحث العلمي، بل وفي الحالات جميعها التي لا يتركز فيها اهتمام الفاحص على تفسير الدرجة الخام التي يحصل عليها الفرد المفحوص وتعرّف وضعه النسبي ضمن المجموعة (مخائيل، 1996).

الفصل الثالث

بناء الاستبانة وتوفير مستلزمات صلاحها



الفصل الثالث

بناء الاستبانة وتوفير مستلزمات صلاحها

مقدّمة:

الاستبانة هي إحدى الأدوات الهامة في عملية البحث والقياس النفسي والتربوي يستخدمها المربّون على نطاق واسع بهدف جمع المعلومات والبيانات بطريقة منظمّة ودراسة المواقف والميول والآراء ومظاهر التكيّف الشخصي والاجتماعي. ولقد عرفت الاستبانة كطريقة في طرح الأسئلة منذ زمن بعيد، ويعود الفضل إلى علماء الاجتماع في القرن العشرين من أمثال هايمان وستوفر في الإلحاح على أهميتها وضرورة إخضاعها للتحليل المنطقي والإحصائي وتقنينها واستخدامها كأداة علمية لجمع المعلومات. وتفيد الاستبانة في دراسة مختلف جوانب العملية التربوية، وتحتل بالتالي مكانة خاصة في الدراسات والبحوث التربوية بأنواعها وخاصة منها البحث الوصفي.

وقد أشار لازارسفيلد وسيبر إلى أن ثلث البحوث التربوية المنشورة عام 1963 في اربعين مجلّة تربوية اعتمدت على الاستبانة كأداة بحث (LAZARSFELD AND SIEBER, 1964). وذكر فان دالن أن "الاستبانة في بعض الدراسات، أو في مراحل معيّنة منها، هي الأداة العملية الوحيدة لاستخلاص المعلومات والبيانات المطلوبة لتأكيد الفرضية المطروحة أو رفضها. (VANDALEN, 1973, P.324).

وسوف نسعى في هذا الفصل إلى إلقاء الضوء على الاستبانة وموقعها في البحث وفي عملية القياس والتقويم التربوي وفاعليتها كأداة بحث وقياس، كما سنسعى إلى معالجة طائفة واسعة من المسائل المتّصلة بشكلها ونوعها وعملية تصميمها وتوفير مستلزمات صلاحيتها والتثبّت من هذه الصلاحية بما يمكن أن يحقّق بعض الفائدة لمن يعمل على إعداد استبانة تتحقّق فيها الشروط

والمواصفات الفنية اللازمة لهذه الأداة الهامة من أدوات البحث التربوي. ولن يكون من أهداف هذه الدراسة الإحاطة بكافة المسائل التي تطرحها عملية تطوير هذه الأداة وسوف نكتفي بالوقوف عند المسائل الأساسية.

الاستبانة ووزنها كأداة في عملية البحث والقياس والتقويم التربوي:

الاستبانة هي ترجمة للكلمة الإنكليزية QUESTIONNAIRE ، ولها في اللغة العربية ترجمات (أو مقابلات) عديدة هي: الاستبيان والاستقصاء والاستجواب والاستخبار والاستفسار والاستطلاع والاستفتاء. وتشير هذه الترجمات المتعدّدة للمصطلح الأجنبي الواحد إلى عدم اتفاق وجهات نظر الباحثين والمترجمين العرب حول المصطلح العربي الأدق والأكثر تعبيراً عن المصطلح الأجنبي، كما يمكن أن تدل على غنى وتنوّع المفردات العربية المقابلة لهذا المصطلح. وفي كل الأحوال تستخدم هذه الكلمات جميعها للدلالة على وسيلة واحدة لجمع المعلومات والبيانات قوامها الاعتماد على مجموعة من الأسئلة تدور حول موضوع معيّن وتقدّم إلى عيّنة من الأفراد للإجابة عنها.

وتعتمد الاستبانة على التقرر الذاتي اللفظي وتوصف بأنها "مقابلة جماعية AMass interview أو مقابلة بالكتابة، وتكون من هذه الناحية أقرب إلى الاختبار الجمعي على حين أن المقابلة أقرب إلى الاختبار الفردي" .(Shoflied, 1972, p.184)

تتفاوت وجهات نظر الباحثين بصدد الاستبانة ومكانتها وقيمتها كأداة في البحث التربوي. فمن الباحثين من يرى أنّها أداة بحث فعّالة واقتصادية وقادرة على تغطية مختلف جوانب العملية التربوية ودراسة مشكلاتها العديدة المتشعّبة (ماتيوشكين، 1981، ليرنر، 1983). ومنهم من يرى أنّها أداة بحث "شعبية" لان كل فرد يعتقد أن بإمكانه أن يطرح مجموعة من الأسئلة حول موضوع معيّن (كابانوفا، 1982). ومنهم من يميل إلى الحط من قدرها والتشكيك بفاعليتها

نظراً لأنّها تقتصر على الوصف ويشيع استخدامها بصورة خاصة في البحث الوصفي الذي ينظر إليه على أنه في المرتبة الدنيا من مراتب البحث. هذا بالإضافة إلى انّها يمكن أن تقدّم معلومات وبيانات غير موثوقة بصورة كافية (Borg, 1979, P.284). والواقع أن الاستبانة كغيرها من أدوات البحث تجمع العديد من المزايا كما تعاني من العديد من نقاط الضعف. ولكل أداة ميزاتها الخاصة وافضليتها على غيرها من الأدوات في نطاق الأغراض الخاصة المرسومة للدراسة وطبيعة هذه الدراسة ومجالها ومستلزماتها. ومن الصعب الحكم على قيمة الاستبانة ووزنها ضمن أدوات البحث بمجموعها في معزل عن الاعتبارات السابقة. فالاستبانة بوصفها أداة لجمع المعلومات والبيانات حول الظاهرات التربوية المختلفة ودراسة المواقف والميول والآراء تغطي مجالات واسعة وعديدة ومتنوّعة يصعب أن تطالها بقية الأدوات. والواقع أن الاستبانة تستعمل على نطاق واسع من قبل المربّين للحصول على معلومات حول الوقائع الماضية والحاضرة والمتوقّعة. وإذا أحسن إعدادها وتوافرت فيها شروط الصلاحية (الصدق والثّبات والكفاية) فإنّها تتدارك الكثير من العيوب وترسّخ الكثير من المزايا. وقد يعود السّبب في الاستهانة بهذه الأداة إلى الانخداع بالبساطة الظاهرية والاعتقاد بيسر وسهولة عملية بنائها والتي لا تتعدّى طرح مجموعة من الأسئلة حول موضوع معيّن. وبخلاف ما يظنّه الكثيرون تعد عملية طرح الأسئلة بحد ذاتها عملية بالغة التعقيد والحساسية وتمثّل إحدى الصعوبات الكبرى في تصميم هذه الأداة وتطويرها. وكما يؤكد شوفيلد " فإن أي إنسان يستطيع أن يطرح الأسئلة ولكن لا يستطيع أي إنسان أي يبني استبانة جيدة " (Shofield, 1972, p.192) وقد أشار أوبنهايم في هذا الصدد إلى أن "طرح السؤال على الناس هو أشبه ما يكون بمحاولة الإمساك بسمكة هاربة برمي أنواع من "الطعم" إلى أعماق مختلفة دون معرفة ما يجري تحت السطح" (Oppemheim, 1966, p.42) كما أكّد فان دالن أن ".. طرح الأسئلة التي ستتكوّن منها الاستبانة والتي تؤدّي إلى بيانات ومعلومات دقيقة وضرورية لاختبار الفرضية المطروحة ليس بالمهمّة السّهلة على الباحث . وعموماً تعد الاستبانة أداة "سهلة ممتنعة" وتتطلّب عملية تصميمها

وتضمينها الأسئلة الملائمة وتحقيق شروط صلاحيتها جهداً كبيراً وعملاً دؤوباً متأنياً. فالأسئلة التي ستتكوّن منها الاستبانة يجب أن توضع بحيث تكون عيّنة من المثيرات التي توجّه لاستدعاء عيّنة من الاستجابات. ويجب أن تكون هذه العيّنة صادقة وممثّلة للمجتمع الأصلي من المثيرات والاستجابات لكي تعبّر عن السّمة أو الظاهرة أو المتغيّرات موضوع الدراسة بأمانة وصدق. وغني عن البيان أن الاستبانة تخضع في سياق عملية تصميمها وحتى تخرج بصورتها النهائية وتصبح أداة جاهزة وصالحة لمؤثّرات مختلفة ومتعدّدة المصادر والاتجاهات ممّا يزيد من صعوبة العمل بها لتكون ميزاناً "حسّاساً" ودقيقاً، إلاّ أن هذا لا ينفي " أن الكثير من نقاط الضعف التي تعاني منها هذه الأداة يمكن تلافيها إذا أعدّت بعناية وتمّ تطبيقها على مبحوثين مؤهّلين " (Vam Dalen, 1973, p.327). والواقع أن الكثير من البحوث التربوية التي تستخدم الاستبانة كأداة وحيدة، أو كواحدة من الأدوات (ولا يستثنى منها بعض البحوث المعدّة لنيل الدرجات العلمية العالية في التربية وعلم النفس) تعاني من عدم مراعاة الشروط الضرورية لتصميم هذه الأداة ومن غياب مؤشّرات صلاحيتها بصورة كلّية أو جزئية ممّا ينعكس مباشرة على نتائج تلك البحوث ويقلّل من قيمتها المنهجية والعلمية. ولعلّ السّبب الأساسي الكامن وراء ذلك هو الاستهانة بعملية تطوير هذه الأداة وتأمين متطلّبات صلاحيتها ممّا يلحق ضرراً كبيراً بها ويحد من فاعليتها كأداة بحث ويشكّك بقيمتها. وقد يكون من الأجدى للباحث، وخاصة الباحث المبتدئ، أن يستعين باستبانة جاهزة تتوافر فيها شروط الصلاحية، على أن يتم التأكّد من هذه الصلاحية واختبارها تجريبياً إذا اقتضى الأمر، بدلاً من أن يقوم بنفسه بوضع استبانة لا تراعي تلك الشروط أو تراعيها بصورة غير كافية ظنّاً منه أن عملية تصميم الاستبانة هي عملية يسيرة وسهلة المنال. فإذا رغب الباحث في تصميم الاستبانة بنفسه ترتّب عليه أن يتزوّد بالمعارف والمهارات الضرورية اللازمة لبناء هذه الأداة. وما من شك في أن الاطلاع على استمارات الاستبانات السّابقة المشابهة يفيد في استدراك الكثير من نقاط الضعف كما يوفّر الكثير من الوقت والجهد على الباحث. ويمكن إذا اقتضى الأمر استعارة أسئلة من استمارات سابقة

على ان تتم الإشارة إليها وإخضاعها كبقية الأسئلة للدراسة والتجريب. ولابد للباحث قبل أن يبدأ في تصميم الاستبانة ان يقرّر فرضيته (أو فرضياته) وأن يحدّد أهداف بحثه بدقّة لكي يتمكّن في ضوئها من تحديد الأغراض الخاصة بالاستبانة والمجال الخاص الذي تسعى إلى تغطيته.

شكل الاستبانة ونوعها:

تستلزم عملية تصميم الاستبانة وإعداد الأسئلة اللازمة لها أن يتّخذ الباحث قراراً بصدد شكلها (الشكل المغلق أو المفتوح أو المغلق – المفتوح)، وأن يقرّر ما إذا كان سيطبّقها على المبحوثين (المستجيبين) مباشرةً أو أنه سيرسلها لهم بالبريد (أو سينشرها على صفحات الصحف والمجلاّت أو على شاشة التلفزيون أو عن طريق الإذاعة أو الهاتف كما يحدث في حالات معيّنة). ولكل من هذه الأشكال والأنواع مزاياه وعيوبه. ويتحدّد اختيار الباحث لشكل الاستبانة ونوعها في ضوء مجموعة من المتغيّرات منها طبيعة الدراسة ومستلزماتها وأهدافها والأغراض الخاصة بالاستبانة، وإمكانات الباحث الواقعية، ونوع العيّنة المختارة وحجمها وتوزّعها الجغرافي... الخ.

تستخدم الاستبانة (غير البريدية) والتي يتولّى الباحث أو أحد مساعديه تطبيقها على المبحوثين في الحالات التي يمكن فيها جمع المبحوثين في مكان واحد وتوزيع الاستمارات عليهم كما هو الحال بالنسبة للطلبة في المدارس. ويشغل هذا النوع حيّزاً هاماً في البحوث التربوية ويتمتّع بمزايا هامة منها أنه يؤمّن الاتصال الحي والمباشر بين الباحث والمبحوثين ممّا يمكّن الباحث من شرح أهداف الدراسة ومغزاها وفائدتها وتوضيح النقاط الخافية عليهم، كما يمكّنه من حفز المبحوثين وتحميسهم واستثارة الدافعية لديهم للإجابة بعناية وأمانة، هذا بالإضافة إلى انه يتيح تفحّص استمارات الاستبانة للتأكّد من تعبئتها تماماً من قبل المبحوثين كافة. وتتميّز الاستبانة غير البريدية بأنّها قليلة التكاليف وأن نسبة الرّدود عليها تكون عالية جداً وأن وجود الباحث بنفسه (أو من ينوب عنه)

وتأكيده سرية البيانات يخفّف من قلق المبحوثين وشكوكهم، ممّا يؤدّي إلى رفع مستوى الصدق في استجابات المبحوثين (صدق الاستجابة) ويرفع بالتالي من مستوى صدق الاستبانة ككل. وتتطلّب هذه الطريقة تواجد شخصين أو أكثر للإشراف على التطبيق (بحسب حجم المجموعة ومستوى ثقافتها) وتقديم العون اللازم بطريقة غير موحية، وتفحص الاستمارات للتأكّد من تعبئتها تماماً (Oppenheim, 1966).

وتتفوّق الاستبانة البريدية على غير البريدية في إمكان تطبيقها على عيّنة كبيرة من الأفراد ومنتشرة في رقعة جغرافية واسعة وعلى مسافات متباعدة. وهي اقتصادية نسبياً وتنحصر تكاليفها في الطباعة والإرسال البريدي ويمكن تعميمها على إعداد كبيرة مع زيادة متواضعة في التكاليف. ثم إن عملية المعالجة والتحليل لنتائج هذه الاستبانة يسيرة وسهلة بالمقارنة مع طريقة المقابلة ولا تحتاج إلى طاقم مدرّب من الأشخاص العاملين في الميدان (Oppenheim, 1966). ولكن يعاب عليها انّها لا يمكن أن تطبّق على أشخاص محدودي الذّكاء أو من درجة ثقافة متدنّية، كما يعاب عليها أنّها تؤدي إلى اعتماد ردود متواضعة وارتفاع نسبة الفاقد. وقد ظهر أنه في حال تطبيق هذه الاستبانة على عيّنة من الأفراد غير المهتمّين بموضوعها مباشرة فإن نسبة الرّدود تتراوح بين 40 – 60٪ وأن هذه النسبة نادراً ما تتجاوز 80٪ حتى إذا كان للمبحوثين مصلحة أو اهتمام مباشريها (Cohen, 1980, p.88). وقد أجريت دراسات عديدة للكشف عن الفروق العقلية والشخصية القائمة بين المستجيبين وغير المستجيبين وأظهرت هذه الدراسات عدم وجود فروق بين المستجيبين وغير المستجيبين في أبعاد الشخصية الهامة. ويميل غير المستجيبين إلى أن يحقّقوا نجاحاً أكاديمياً أقل من المستجيبين، ويؤدّي انخفاض نسبة الرّدود إلى انحياز العيّنة نظراً لأن حالات عدم الاستجابة لا تخضع للتعيين العشوائي ولها محدّداتها الخاصة التي قد تختلف من مسح لآخر. ويصعب التغلّب على معضلة الفاقد وانحياز العيّنة دون معرفة طبيعة الانحياز. ويرى يورج ان الطريقة الأمثل

لمواجهة هذه المعضلة هي اختيار عدد من الحالات بطريقة عشوائية من فئة غير المستجيبين وإجراء مقابلات معهم. فإذا كانت هذه الفئة منتشرة على رقعة جغرافية فيمكن إجراء المقابلة مع الأقرب منهم إلى مكان وجود الباحث. وبعد الحصول على البيانات المطلوبة من هؤلاء لابد من مقارنة استجاباتهم لكل بند من بنود الاستبانة مع استجابات أولئك الذين ملأوها وأعادوها «فإذا لم تظهر فروق أو ظهرت فروق طفيفة يمكن الافتراض أن عيّنة المستجيبين ليست عيّنة منحازة، وإذا ظهرت فروق كبيرة فلا بد من الإشارة إليها ومناقشة دلالتها عند تقرير النتائج. ومن المفيد بهدف خفض نسبة الفاقد أن ترافق الاستمارة البريدية رسالة موجزة تشرح أغراض الدراسة وأهميتها وتحتوي على أسباب ومسوغات مقنعة لتعبئة الاستمارة وإعادتها. كما أن من المفيد اللجوء إلى رسائل التذكير والمكالمات الهاتفية والبرقيات (Cohen,1980, Borg, 1979). ولا يوجد اتفاق بين الباحثين حول الحد الأدنى المقبول لنسبة الردود والحد الأقصى المسموح به لنسبة الفاقد. ومنهم من يرى أن نسبة الفاقد لا يجوز أن تتعدّى 50٪ (التير، بلا تاريخ، ص135). والمهم في كل الأحوال هو أن تنخفض نسبة الفاقد إلى أدنى درجة كي لا تؤدي إلى إحداث نقص كبير في عدد أفراد العيّنة الكلية، كما أن من المهم أن تتوزّع هذه النسبة توزّعاً عشوائياً كي لا تؤثّر سلباً في صدق العيّنة وتمثيلها. فإذا جاء هذا التوزيع منحازاً ترتّب على الباحث إجراء بعض التصحيحات اللازمة. وغني عن البيان أن على الباحث أن يتوقّع مسبقاً امتناع عدد من المبحوثين عن الإجابة وان يحسب حساباً لذلك عند تحديد عدد أفراد العيّنة الذين سترسل لهم الاستبانة.

وتجدر الإشارة إلى أنه ظهر حديثاً في بعض البلدان المتقدّمة ما يعرف باسم "الاستبانة الهاتفية" وتعتمد هذه الطريقة في إجراء الاستبانة وتطبيقها على المبحوثين على أن يقوم الباحث أو أحد مساعديه بالاتصال هاتفياً بأفراد العيّنة وقراءة الأسئلة عليهم وتسجيل إجاباتهم عبر الهاتف. ومن مزايا الاستبانة الهاتفية أنها تؤمّن نوعاً من الاتصال المباشر بين الباحث والمبحوثين ممّا يعطي

المجال للباحث لشرح أهداف الدراسة واستثارة الدافعية لديهم للإجابة، إلاّ أن هذا النوع لا يصلح إلاّ في البلاد التي ينتشر فيها الهاتف انتشاراً واسعاً، كما لا يصلح إلاّ في حالة الاستبانة القصيرة حيث لا يرغب أكثر الناس في الإجابة عن أسئلة كثيرة عبر الهاتف. ومن الصعب تطبيق هذه الاستبانة على عيّنة كبيرة ومنتشرة انتشاراً واسعاً، ويمكن تطبيقها في حالات خاصة وعلى عيّنة صغيرة نسبياً، كما يمكن الإفادة منها في الدراسة الاستطلاعية.

إن اختيار الشكل الملائم للاستبانة (الشكل المغلق أو المفتوح أو الشكل المغلق – المفتوح) يخضع لمجموعة من المحدّدات أبرزها طبيعة الموقف الذي ستطبّق فيه الاستبانة وخصوصيته بالإضافة للأغراض الخاصة المرسومة للاستبانة. ومن الصعب الحكم على أحد الشكلين بأنّه مناسب أو غير مناسب في كل الحالات فلكل منهما مزاياه التي تتأكّد في مواقف وحالات معيّنة ولكل منهما عيوبه التي تبرز في مواقف وحالات أخرى. وعموماً يتفوّق الشكل المفتوح على الشكل المغلق في مجال الكشف عن دوافع المبحوثين ومواقفهم وفي إتاحة الفرصة لهم لاستخدام تعابيرهم الخاصة في الإجابة وتقديم استجابات "عفوية " لا تتقيّد بالبدائل التي يتضمّنها الشكل المغلق. فإذا كان من الصعب على ا لباحث أن يتوقّع إجابات المبحوثين، وإذا رغب في أن يتحدّث المستجيب بحرية بحيث يتمكّن فيما بعد من تعرّف آرائه الخاصة والأسباب الكامنة وراءها، وتحليل أقواله للحصول على تفصيلات مفيدة فما عليه إلاّ أن يستخدم الشكل المفتوح (Shofield, 1972). (Van Dalen, 1973) إلاّ أن الشكل المفتوح يعاني من عيوب كبيرة منها أن المبحوثين (المستجيبين) يفتقرون إلى الإشارات أو التلميحات التي يمكن أن توجّه تفكيرهم وقد لا يفهمون المقصود من السؤال (الأسئلة) فيحذفون عن غير قصد معلومات هامة وخاصة إذا كانوا من مستويات تعليمية متدنّية. كما أن عملية تفريغ وجدولة الإجابات عن هذا الشكل والتي غالباً ما تأتي متنوّعة تنوّعاً واسعاً، ليست بالعملية السّهلة وتستلزم وقتاً وجهداً كبيرين.

ولقد اشار أوبنهايم بهذا الصّدد إلى أن «... من السّهل طرح الأسئلة المفتوحة ولكن من الصّعب الإجابة عنها، والأصعب هو تحليلها».

(Oppenheim, 1966, p.34)

ينطوي الشّكل المغلق أو المقيّد على احتمالات أو بدائل محدّدة يضعها الباحث للكشف عن النواحي التي يرغب في دراستها لدى المبحوثين و«توجيه» تفكيرهم نحو تلك النواحي دون غيرها. ويتجنّب الباحث باستخدامه هذا الشّكل احتمال عدم فهم الأسئلة أو سوء فهمها والاستطراد المحتمل في الإجابة عن الأسئلة المفتوحة، ومن هذه الزاوية يتمتّع بميزة هامة وهي أنه يضم أسئلة وبدائل محدّدة تقنّن طريقة الإجابة كما تقنّن طريقة تحليل الإجابات. ويفترض أن تؤدّي هذه الطريقة في طرح الأسئلة إلى «إنقاص عدد الذين يمتنعون عن الإجابة لعدم فهم السؤال وإنقاص عدد من يسيء فهم الأسئلة أو المقصود منها فتكون إجاباتهم غير دقيقة ولا معبّرة عن آرائهم» (Shofield, 1972, p.187) ويسهّل الشّكل المغلق عملية التفريغ والجدولة والتحليل ويحقّق بذلك إلى درجة معيّنة شرطاً هاماً من شروط صلاحية الأداة وهو الاقتصاد في الوقت والجهد. والواقع أن هذا الشّكل هو الشّكل الأكثر انتشاراً بين الباحثين على الرغم من بعض الانتقادات الموجّهة له والتي تتركّز في أنه قد يضم أسئلة «مشحونة» وموحية ويمكن أن يؤدّي إلى الخسارة في عفوية المستجيب وتعبيريته، وأنه غالباً ما يعجز عن الكشف عن دوافع المستجيب والأسباب الكامنة وراء اختياره إجابة أو إجابات معيّنة دون غيرها، وقد يضطر المستجيب إلى إعطاء إجابة لا تعبّر عن افكاره بدقّة طالما أنه مقيّد بمجموعة البدائل الجاهزة أو «أن يعطي إجابة تنسجم مع رغبات الباحث وتؤيّدها» (Van Dalen, 1973, p.326). ومن هنا لجأ الباحثون إلى استخدام الشّكل المغلق — المفتوح ورأى بعضهم أن تترك صفحة بيضاء خاصة بالأسئلة المفتوحة. كما رأى بعضهم استخدام الأسئلة المفتوحة أو بعضها في مرحلة العمل الاستطلاعي بهدف تحويلها فيما بعد إلى أسئلة مغلقة. وتحقّق الأسئلة المفتوحة في هذه الحالة الأخيرة فائدة كبيرة فهي تفسح المجال لإسهام

أفراد العيّنة الاستطلاعية في اقتراح البدائل، وفي هذا إثراء للاستبانة من حيث أن المبحوثين أنفسهم يسهمون في عملية بنائها وتطويرها (يرماكوفا، 1979، ماتيوشكين، 1981).

تحديد الأغراض الخاصة للاستبانة:

لقد تمّت الإشارة إلى أن عملية تحديد نوع الاستبانة المراد تصميمها وشكلها تخضع لمجموعة من المتغيّرات أبرزها الأغراض الخاصة بهذه الأداة. ومن المفيد الإشارة ها هنا إلى أن تحديد الأغراض الخاصة بهذه الأداة لابد أن يسبق الخطوة العملية الأولى في تصميمها وهي إعداد الأسئلة في صورتها الأولية، ويسهم تحديد مشكلة البحث وفرضياته وأهدافه إسهاماً كبراً في تحديد الأغراض الخاصة بالاستبانة والمجالات التي يسعى الباحث إلى تغطيتها من خلال الأسئلة التي سيعمل على إعدادها. ويرى بورج أن من الصعب على الباحث أن يوضّح الأسباب الكامنة وراء اختياره وطرحه أسئلة معيّنة دون غيرها وكيف سيقوم بتحليل استجابات المبحوثين دون أن يملك فكرة واضحة ومحدّدة عن مشكلة البحث وأهدافه. كما أن «من الصعب دون تحديد أغراض الاستبانة مسبقاً اتخاذ قرارات صائبة حول اصطفاء العيّنة وتصميم الاستبانة وطرائق تحليل البيانات» (Borg,1979, p.293). وتبعاً لذلك يترتّب على الباحث أن يعمل على تحديد مشكلة البحث وبيان عناصرها الرئيسية كما يترتّب عليه أن يعمل على تحديد المجالات الرئيسة التي ستدور حولها الأسئلة من خلال الفروض التي تطرحها الدراسة والمتغيّرات التي تتّجه تلك الفروض إلى كشف الصّلة بينها. ومن المفيد بعد تحديد المجالات الرئيسة للاستبانة (وتفريعها إلى مجالات فرعية إذا تطلّب الأمر) وضع قائمة بالنقاط التي يتضمّنها كل مجال وتحديد عدد الأسئلة اللازمة لكل نقطة من تلك النقاط. وليس من الضروري أن يكون عدد الأسئلة متساوياً في المجالات أو النقاط كافة.

ولابد من أن يتناسب عدد الأسئلة اللازمة لكل مجال ولكل نقطة من النقاط التي يتضمّنها مع أهمية هذا المجال وأهمية النقاط التي يتضمّنها والوزن «النسبي» لكل منها (حسن، 1982، ص 350).

الصورة الأولية للاستبانة:

يتطلّب تصميم الاستبانة بصورتها الأولية إعداد مجموعة كبيرة من الأسئلة حول كل مجال من مجالات الاستبانة والنقاط التي يضمّها. ويفيد الإكثار في عدد الأسئلة عن الحد المطلوب في المفاضلة بينها واختيار الأصلح منها لأغراض الدراسة و«غربلتها» في ضوء نتائج التطبيق التجريبي. وليس ثمّة اتفاق بين الباحثين حول النسبة المئوية للأسئلة الزائدة عن العدد المطلوب، ويترتّب على واضع الاستبانة نفسه أن يقدّر هذه النسبة. والواقع أن عملية إعداد الصورة الأولية للاستبانة بعناصرها الكثيرة المتعدّدة ليست بالأمر السّهل وتتطلّب مهارات خاصة لدى مصمّم الاستبانة وقد تبرز الكثير من شخصيته وكفاءته العلمية. ولا تكفي المهارات التعبيرية والقدرة على الصياغة اللغوية السليمة وحدها لطرح الأسئلة ولابد من توافر مهارات وشروط أخرى.

اسأل؟ وكيف؟ ولماذا؟ وكيف سأقوم بتحليل الاستجابات؟ وقد يظنّ البعض أن عملية طرح الأسئلة هي عملية سهلة ولكنّها ليست كذلك في الواقع فما نسأله هو واضح لنا لمعرفتنا بالجواب وقد لا يكون واضحاً بالنسبة للمبحوث (Oppenheim,1966, p.39) . ويشير فان دالن في هذا الصدد إلى أن «صياغة الأسئلة بهدف الحصول على إجابات صادقة هي فن قائم بذاته».

(Van Dalen, 1973, p.327)

ومن الباحثين من يرى ضرورة مراعاة التسلسل الزمني في صياغة الأسئلة (من الماضي إلى الحاضر فالمستقبل) والانتقال من المعلوم إلى المجهول ومن المحسوس والبسيط إلى المجرّد والأكثر تعقيداً (Shofild, 1972, p.189).

ومنهم من يرى أن توضّع الأسئلة وفق الترتيب المنطقي مع مراعاة الترتيب «السيكولوجي» وبحيث تسبق الأسئلة السّهلة والمشوّقة والمحايدة الأسئلة الأكثر صعوبة والأقل تشويقاً والمتّصلة بالنواحي الشخصية.

(Vam Dalen, 1973, p.328)

كما أن من الباحثين من يعتقد بوجود صعوبة في تقديم مبادئ وأسس ثابتة لترتيب الأسئلة ويرى أن يتم اختيار الترتيب الملائم للأسئلة «في ضوء مشكلات المسح الخاصة ونتائج العمل الاستطلاعي».

(Oppenheim, 1966, p.40)

وعموماً لابد أن تتوافر في الأسئلة صفات الوضوح والتحديد وأن يراعى في ترتيبها التسلسل المنطقي دون إهمال الترتيب «السيكولوجي» ومتطلّباته وان تتناسب مع سن المستجيب ودرجة تعليمه ومستوى فهمه، وتكون شاملة بحيث تغطّي تفاصيل البيانات المطلوبة. ويجب أن ينظر إلى السؤال على انه مثير له استجابة ثابتة نسبياً وإلا فلا معنى لطرحه لأن الاستجابة (الاستجابات) سوف تتذبذب وتضعف مستوى الثبات (الموثوقية). كما يجب أن تصاغ الأسئلة بحيث «تكون الاستجابة على السؤال على أساس استجابتهم للضغوط الاجتماعية في الموقف، ويشار إلى هذه الإجابة عادة على أنها تعميمات جامدة» (جابر،1987، ص252). وعلى هذا النحو لابد أن تأخذ الأسئلة صيغاً واضحة المعنى والمبنى وان يتجنّب الباحث السؤال الذي يضم أكثر من فكرة أو نقطة واحدة (السؤال المركّب) فمثل هذا السؤال يمكن تجزئته بحيث يصبح سؤالين فرعيين أو أكثر. كما لابد من تجنّب الأسئلة التي قد تنطوي على أكثر من معنى واحد وتحتمل تفسيرات عديدة، والأسئلة غير المحايدة أو الموجبة أي التي توحي بجواب واحد مقبول، من مثل: «معظم الناس ينظّفون أسنانهم يومياً. الا تفعل أنت ذلك؟ أو «معظم الناس هذه الأيام يعتقدون بـ ... ألا تعتقد أنت بذلك أيضاً؟ «ومن الأفضل تجنّب الأسئلة التي تأخذ صيغة النفي بالإضافة إلى الأسئلة

التي تحتاج الإجابة عنها إلى ثقافة رفيعة، والأسئلة المثيرة والمزعجة من مثل: «هل قرأت صحفاً أو مجلاّت من أي نوع كان في أثناء حياتك المهنية كمعلّم أو هل شاركت في دورة تدريبية من أي نوع كان في أثناء حياتك المهنية كمعلّم؟».

ومن المناسب في حالات أن تحتل أسئلة البيانات الشخصية من مثل الأسئلة التي تتعلّق بالعمر والجنس والوضع العائلي.. الخ مكانها في مواضع متأخّرة من الاستبانة لكي يستجيب لها المبحوث بصورة إيجابية وبعد أن يكون قد اقتنع بأهمية البحث. ولهذه الأسئلة أهمية خاصة في تصنيف العيّنة حسب المتغيّرات المطروحة للدراسة. (Oppenheim, 1966,p.58)

ومن الأمور التي يجب أن يأخذها الباحث بالحسبان ألاّ تكون الأسئلة طويلة وعبئاً على المبحوث من ناحية الوقت. فالزمن الطويل قد يؤدي إلى السأم والتعب ويؤثّر في جدية المستجيب ويضعف اهتمامه. ولا يوجد قاعدة عامة بصدد الطول المناسب للاستبانة فثمّة فروق كبيرة بين الناس في سرعة الفهم وتحمّل التعب وسرعة الاستجابة. وعموماً فإن احتمال إجابة أفراد العيّنة عن الاستبانة يزداد كلّما قصرت، كما ان الاستبانة التي يحتاج الشخص المتوسّط لتعبئتها إلى أكثر من ساعة تعدّ استبانة طويلة (التير، بلا تاريخ، سيفيكينا، 1986). ومن هذه الزاوية لابد أن يؤخذ الوقت بالحسبان ومن خلال مراعاة الطبيعة الخاصة للعيّنة سواء من ناحية العمر أو المستوى التعليمي أو غير ذلك على أن تكون الاستبانة من الطول بحيث تغطّي جميع البيانات الضرورية للدراسة في الوقت نفسه. ومن الأمور الهامة في هذا السياق استثارة الدافعية لدى المبحوثين للإجابة بعناية وصدق عن كافة الأسئلة. وثمّة عوامل عديدة تؤدّي إلى رفع مستوى الدافعية والاهتمام لدى المبحوثين منها تفهّم أغراض الدراسة وقيمتها والاقتناع بأهميّتها وجدواها، والصياغة الواضحة والمحدّدة للأسئلة وتعليمات الإجابة، والابتعاد ما أمكن عن الأسئلة الشخصية المحرجة وتجنّب الأسئلة الاستفزازية والمثيرة للقلق (كالأسئلة التي تتعلّق بالدين أو الطائفة أو الدّخل غير المشروع... الخ).

وكثيراً ما تنص التعليمات على أنّ الإجابة لن يطلع عليها أحد بل هي لأغراض البحث فقط. ويطلب عادة عدم كتابة الاسم إلاّ في حالات خاصة وعندما يرغب الباحث في إعادة تطبيق الاستبانة لمعرفة التغيّر (والثبات) في نظرات المبحوثين ومواقفهم (Shofield, 1972). ومن المفيد بصورة عامة الحفاظ على سرية الاسم بهدف الحصول على استجابات صريحة وكاشفة، وفي الحالات التي تتطلّب ذكر الاسم يمكن إعطاء المستجيبين أرقاماً بدلاً من مطالبتهم بذكر أسمائهم. كما أن من المفيد بهدف ترغيب المبحوثين أن تكون الاستبانة منظّمة وحسنة التبويب وجذّابة من ناحية الشّكل والإخراج.

تعليمات الاستبانة:

للتعليمات المرافقة للاستبانة أهمية خاصة من حيث أنّها ترشد المبحوث وتبيّن له ما يجب عليه أن يفعله وتؤثّر بالتالي تأثيراً كبيراً في صدق الاستجابة. ونظراً لعدم وجود احتكاك شخصي مباشر بين الباحث والمبحوث، وخاصة في الاستبانة البريدية، يترتّب على الباحث أن يولي عناية فائقة بصياغة التعليمات، وأن يتنبّأ باحتياجات المبحوث ويقدّم له التوجيهات اللازمة لفهم طريقة الإجابة. ويجب أن تتوافر في التعليمات شروط البساطة والوضوح والتحديد وأن تكون مباشرة ووافية بالغرض دونما تطويل زائد أو إيجاز غامض. ويفضّل أن تكتب بخط عريض أو يوضع تحتها خط. وتخضع التعليمات للتعديل والإضافة (والحذف أحيانا) في مراحل العمل جميعاً كما تخضع الأسئلة ذاتها. ويتم من خلال التجربة الاستطلاعية بوجه خاص واللقاء المباشر أو «المقابلات» التي يجريها الباحث مع أفراد العيّنة الاستطلاعية تعرّف الصعوبات «الواقعية» التي يواجهها المبحوثون في تعبئة استمارة الاستبانة، وفي ضوء ذلك يتم إجراء التعديلات اللازمة في التعليمات وصياغتها. ولا شك أن لعدم وضوح التعليمات وسوء فهم الأسئلة الناجم عنه أثر سلبي مباشر في صلاحية الاستبانة من حيث أنه يلحق ضرراً كبيراً بصدقها وموثوقيتها (ثباتها). وقد يركّز الباحث اهتمامه بالأسئلة ولا يولي اهتماماً كافياً بتعليمات الإجابة عنها ممّا يسيء إلى الأسئلة ذاتها وقد

يؤدّي إلى سوء فهمها ويدعو بالتالي إلى التشكيك بصحّة الإجابات. ومن هنا لابد من التنبّه لأهمية التعليمات وشمولها وأهمية صوغها في عبارات واضحة ومحدّدة وتعديلها وتنقيحها في المراحل اللاحقة حسب اللزوم.

التحليل المنطقي:

بعد إعداد الأسئلة وتعليمات الإجابة عنها في صورتها الأولية لابد من مراجعتها بصورة متأنّية وإخضاعها للتحليل المنطقي بهدف التحقّق من ملاءمتها للأغراض الخاصة بالاستبانة و«تمثيلها» للمجالات التي تسعى إلى تغطيتها، ومن وضوحها وتسلسلها المنطقي. ومن الجدير بالإشارة هنا أن التحليل العقلي المنطقي لهذه الأداة ليس عملاً متمّماً لها أو لاحقاً بها بل هو عمل مستمر و«مواكب» لعملية تصميمها، ويسبق ظهورها في صورتها الأولية ولكن تكون له في هذه المرحلة بالذات مكانة خاصة. ومن الضروري بهدف المزيد من التحليل عرض الأسئلة المقترحة مع تعليماتها على مجموعة من المختصّين والخبراء أو المحكمين) ممّن لهم صلة بموضوع الدراسة ودراية بتطوير الاستبانة. ويسهم إطلاع المختصّين والخبراء على الأغراض الخاصة المرسومة للاستبانة والمجالات التي تسعى إلى تغطيتها والنقاط الخاصة بكل مجال في التحقّق من تمثيل عناصر الاستبانة للبيانات المطلوبة وملاءمة الأسئلة المطروحة لكل نقطة في مجال ممّا يسهم في تأسيس صدق المحتوى أو الصدق المنطقي للاستبانة . والملاحظات التي يقدّمها أولئك المختصّون والنقاط المشتركة بينهم لابد أن تؤخذ بالحسبان في مراحل العمل اللاحقة وقد يكون لها دور كبير في تطوير الاستبانة ورفع مستوى صلاحيتها إذا أحسن الباحث الإفادة منها. وعموماً فإن للتحلل المنطقي الذي يقوم به الباحث بنفسه كما يقوم به مستعيناً بآراء المختصّين أهمية فائقة من حيث أنه يشكّل الحجر الأساسي في عملية تصميم الاستبانة ولابد أن يواكب هذه العملية حتى نهايتها. وقليلاً ما تولي «الأدبيات» والدراسات الخاصة بتصميم الاستبانة الأهمية اللازمة لهذا النوع من التحليل وغالباً ما تعطي الأولوية للعمل التجريبي والتحليل الإحصائي. والواقع أن من الأهمية بمكان أن يكون للتحليل

المنطقي مكانه الطبيعي وان تعطى له الأولوية طالما أن هذا التحليل هو الأساس الذي ترتكز عليه عملية التصميم برمّتها وأنه يضمن صدق المحتوى أو الصدق المنطقي للاستبانة ويفيد في دراسة جوانب هامة كالترتيب المنطقي (أو السيكولوجي) للأسئلة وصياغتها، هذا بالإضافة إلى انه يسبق العمل التجريبي والتحليل الإحصائي ويحدّد مجال كل منهما واتجاهه وإجراءاته ويكون له دوره في هذه المرحلة ذاتها من مراحل العمل. وعلى هذا فإن تجاهل مصمّم الأداة لهذا النوع من التحلل، أو إجراءه بصورة شكلية أو «روتينية» من خلال تقديم الأسئلة في صورتها الأولية «على عجل» إلى بعض المختصّين دون تقديم بيانات تفصيلية وواضحة حول الأغراض الخاصة المحدّدة لها ومجالاتها الرئيسة والفرعية والأسئلة التي تغطي كلاً من هذه المجالات قد يلحق ضرراً كبيراً بالاستبانة وصدقها، وقد لا يمثّل أكثر من إجراء شكلي أو «استعراضي» يحقّق خدمة واحدة وهي تغطية عيوب الاستبانة أو نقاط ضعفها بالألقاب العلمية أو الأسماء البارزة لأولئك المختصّين.

الدراسة الاستطلاعية:

بعد أن ينتهي الباحث من إعداد الأسئلة بصورتها الأولية وإخضاعها للتحليل المنطقي من قل الباحث نفسه ومجموعة من المحكمين وتعديلها وتنقيحها في ضوء هذا التحليل ينتقل إلى الخطوة التالية في تصميم الاستبانة وهي التجريب الأولي للاستبانة أو ما يعرف بالدراسة الاستطلاعية. ومن نافلة القول أن التحليل المنطقي ونصائح الخبراء ليست بديلاً عن الدراسة الاستطلاعية المنظمّة فهذه الدراسة تأتي مساندة لهذا التحليل ومتمّمة له. ويترتّب على الباحث هنا أن يختار عيّنة من الأشخاص من المجتمع نفسه الذي ستسحب منه عيّنة الدراسة فيما بعد. وقد لا يتجاوز عدد أفراد العيّنة الاستطلاعية 20 أو 30 شخصاً في حالات، وقد يكون من الضروري تجاوز هذا العدد بكثير في حالات أخرى. وتحقّق الدراسة الاستطلاعية فوائد عديدة للباحث حيث يقوم في ضوئها بإعادة صياغة الأسئلة التي يتبيّن أنها تعاني من الغموض،

كما يقوم بإلغاء الأسئلة غير اللازمة، وإضافة أسئلة جديدة أظهرت هذه الدراسة ضرورة تضمينها استمارة الاستبانة. ومن فوائد الدراسة الاستطلاعية أيضاً أنها تمكّن الباحث من «تحويل الأسئلة المفتوحة إلى أسئلة مغلقة بعد أن يكون قد اكتشف من خلال الإجابات المجالات التي يمسّها البحث ونوع الاستجابات اللازمة» (حسن، 1982، ص353). كما أن العمل الاستطلاعي يتيح دراسة كفاية هذه الأداة بجوانبها، بما في ذلك دراسة فعالية طرائق التفريغ والتحليل الكمّي» (Borg, 1979, p.352). ويفيد العمل الاستطلاعي في تحديد الزمن اللازم للإجابة عن الاستبانة وتحديد طولها، كما يفيد في تحديد الصعوبات المتّصلة بصياغة التعليمات وما إذا كانت محدّدة وواضحة وفي مستوى الفهم العام للمبحوثين. هذا بالإضافة إلى أنه يتيح الوقوف عند الأثر الذي يحدثه ترتيب الأسئلة بهدف إعادة تنظيم الاستبانة وإرجاء بعض الأسئلة التي تبدو أنها محرجة إلى أماكن أخرى من الاستمارة ممّا يسهم بالنتيجة في تخفيض نسبة الفاقد (من غير المستجيبين) (Oppenheim, 1966). وعلى الباحث هنا أن يقوم بتسجيل الملاحظات أو التعليقات التي تكون قد أثيرت حول الأسئلة وان يعمل على دراستها بصورة متأنّية مع الوقوف بصورة خاصة عند الأسئلة التي يمتنع الكثيرون عن الإجابة عنها لسبب أو لآخر والأسئلة التي تكثر الإجابات المحايدة عنها (من نوع لا أدري) أو الأسئلة التي قد تؤدّي إلى إجابة واحدة لدى المبحوثين جميعاً. وعموماً تسهم الدراسة الاستطلاعية والتحليل الشامل للأسئلة في ضوئها في تقديم بيانات مفيدة حول مؤشّرات صلاحية الاستبانة بصورة أولية (من حيث الصدق والثبات والكفاية)، وعلى الباحث أن يعمل على توظيف نتائج هذه الدراسة وإجراء التعديلات والتصحيحات اللازمة بما يخدم صلاحيتها ويسهم في رفع مستوى هذه الصلاحية.

التجريب التمهيدي والنهائي:

إن التجريب الأولي (الاستطلاعي) للاستبانة لا يكفي في معظم الأحيان، ويترتّب على الباحث أن يتابع عمله التجريبي بإجراء تجربة أخرى أو سلسلة من التجارب قبل الوصول إلى المرحلة الأخيرة وهي: مرحلة إعداد الاستبانة وإخراجها بصورتها النهائية. فبعد أن تعاد صياغة السؤال أو الأسئلة في ضوء التجربة الاستطلاعية لابد من إخضاعه للتجريب ثانية طالما أن هذه الصياغة الجديدة قد تؤدّي إلى تحيّزات أو صعوبات جديدة. وعندما يحوّل السؤال المفتوح إلى سؤال مغلق لابد من تجريبه في شكله الجديد. ويرى أوبنهايم أن «الأسئلة التي يتم اقتباسها أو استعارتها من استبانات جاهزة، معروفة بصلاحيتها، ولابد أن تخضع بدورها للتجريب لضمان الحصول على نتائج مرضية من العيّنة التي اختارها الباحث (Oppenheim, 1966,p.26). كما يقترح أوبنهايم تجريب الاستبانة على مراحل بحيث يتم تجريب كل فئة من فئات الأسئلة تغطّي مجالاً واحداً بعينه على حدة. وهذا سيجزّئ العمل التجريبي إلى عدد من العمليات الصغيرة، ويعطي الخبرة في التواصل مع المبحوثين، وفي شرح الغرض من الدراسة على نحو أفضل. وبالطبع فإن التجريب المرحلي لا يلغي ضرورة تجريب الاستبانة ككل فيما بعد فلكل منهما فوائده. وفي كل الأحوال لابد أن يولي الباحث اهتماماً خاصاً بالعمل التجريبي والتحلل الإحصائي لنتائجه مع مراعاة شروط ومتطلّبات هذا العمل بمنتهى الدقّة، كما لابد أن ينظر إلى التجريب على أنه المجال الطبيعي «لغربلة الاستبانة وتنقيتها واستكمال أو تأمين مستلزمات صلاحيتها. وتخضع الحاجة لإعادة التجربة لمجموعة من المتغيّرات منها الأغراض الخاصة بالدراسة والنتائج التي أسفرت عنها التجربة الأولى حول الاستبانة وصلاحيتها، والشروط الواقعية المحيطة بالعيّنة، والإمكانات المتاحة للباحث. وعموماً يندر أن تفي التجربة الاستطلاعية وحدها بأغراض التجريب وفوائده وغالباً ما تتطلّب عملية تصميم الاستبانة والوصول بها إلى صورتها النهائية إجراء تجربة لاحقة أو سلسلة من التجارب (التمهيدية والنهائية) بهدف متابعة

عملية «تنقية» الاستبانة ورفع مستوى صلاحيتها. والأمر الذي يجب أن يؤخذ بالحسبان هو أن التجريب ليس غاية بذاته والإكثار منه أو تنويعه ليس دليلاً بحد ذاته على رفع مستوى جودة هذه الأداة. وعلى هذا فإن التجريب لا يجوز أن يكون إجراء «روتينياً» أو شكلياً يقوم به الباحث بهدف الظهور أو «الاستعراض». وتتحدّد قيمة العمل التجريبي برمّته وبكافة أشكاله وبما يتطلّبه من إجراءات المعالجة والتحليل الإحصائي، بما يقدّمه هذا العمل من بيانات «يوظّفها» الباحث فعلاً في صالح عملية التحقّق من صلاحية الاستبانة وفي رفع مستوى هذه الصلاحية إلى الحدود القصوى.

الشكل النهائي للاستبانة:

بانتهاء عملية التجريب والتحليل الإحصائي وبعد أن يكون الباحث قد انتهى من عملية التعديل والتنقيح و«التهذيب» للأسئلة وتعليماتها وتمكّن من اختيار أفضلها وأكثرها تحقيقاً للأغراض المرسومة تبدأ المرحلة الأخيرة في تصميم الاستبانة وهي تنسيق الاستمارة الخاصة بها وإعدادها في شكلها النهائي. وتتطلّب هذه المرحلة من جملة ما تتطلّبه طباعة الاستمارة على وجه واحد فقط لتكون واضحة وسهلة القراءة، وتخصيص المكان الكافي أمام كل سؤال أو تحته للإجابة عنه وإعطاء الأسئلة أرقاماً متسلسلة وتقسيمها إلى مجموعات توضع لها عناوين واضحة. كما تتطلّب وضع علامات مميّزة على الاستمارات لتسهيل التعرّف على كل فئة منها إذا كان الباحث يسعى إلى المقارنة بين إجابات الفئات المختلفة من المبحوثين، وإرسال ظرف معنْوَن عليه طابع بريدي إذا كانت الاستبانة بريدية. وعموماً فإنّ لشكل الاستمارة وإخراجها أثراً لا يستهان به في تعاون المبحوثين ويتوقّع أن ترتفع نسبة الرّدود بصورة ملحوظة مع زيادة الاهتمام بالشّكل والإخراج.

صدق الاستبانة:

لقد تمّت الإشارة في السابق إلى أن عملية تطوير الاستبانة والوصول بها إلى شكلها النهائي تتطلّب التحقّق من صلاحيتها من خلال اختبار صدقها وموثوقيتها (ثباتها) وكفايتها. والواقع أن عملية تصميم الاستبانة لابد أن تواكبها عملية التحقّق من صلاحيتها إذا أريد لها أن تكون أداة بحث علمية ودقيقة وأن تتبعها أيضاً. ويتم التحقّق من الصدق المنطقي للاستبانة (صدق المحتوى) من خلال دراسة عناصر الاستبانة ويضمن هذا الشكل من أشكال الصدق شمول الاستبانة «للعناصر أو أنواع السلوك أو البيانات أو العوامل التي يجري قياسها دون زيادة أو نقصان» (حمدان، 1989، ص145). وقد أشار ايفانز إلى أن هذا الشّكل من أشكال الصدق يتطلّب مراعاة أربع نقاط حيوية وهي: 1 – أن يكون الغرض من الاستبانة واضحاً تماماً في ذهن واضعها وأن يعمل على تعريف المبحوثين بهذا الغرض. 2 – أن يقرّر في ضوء هذا الغرض البيانات المطلوبة بمنتهى الدقّة. 3 – أن يحلّل هذه البيانات إلى عناصرها وأجزائها ويتأكّد من أن الأسئلة الموضوعة تغطّي هذه العناصر والأجزاء و«تتّسق» معها. 4 – أن يضمن أن تكون الأسئلة مصاغة بحيث تحفّز على إعطاء الجواب المطلوب (Evans, 1968, p.41 – 45). وعموماً تفيد عملية اختبار صدق الاستبانة بهذه الطريقة في التثبّت ممّا إذا كانت الأسئلة تسأل فعلاً ما يراد منها أن تسأله، وهي تسبق المرحلة التجريبية وتستمر معها. وتنطوي هذه الطريقة على أهمية قصوى من حيث أنها تكشف عن ملاءمة محتوى الاستبانة وما تضمّه من البنود للموضوع المدروس بمجالاته الرئيسة والفرعية ودرجة تمثيلها له وللمتغيّرات التي يسعى الباحث إلى دراستها وكشف الصّلة بينها. ويؤدّي إهمال هذا الشكل من أشكال الصدق أو عدم الاهتمام به بصورة كافية إلى التشكيك كلّياً أو جزئياً بالاستبانة وصلاحيتها وقيمتها كأداة بحث بغض النظر عن الجهود التي قد بذلها الباحث في مجال العمل التجريبي والتحليل الإحصائي. ومن الطرائق المتّبعة في اختبار صدق الاستبانة اللجوء إلى ما يسمّى بالصدق المحكي حيث تتم

مقارنة نتائج الاستبانة مع نتائج استبانة أخرى تؤخذ كمحك وتقيس الشيء أو (المتغيّر) نفسه الذي تقيسه الاستبانة. ويتميّز الصدق المحكي عن الصدق المنطقي أو الصدق عن طريق المحكمين في أنه يعبّر عن صفة الصدق بلغة الكم من خلال معامل الصدق (وهو معامل ارتباط بطبيعة الحال)، ويعاب عليه إمكان التّشكيك في صدق الأداة ذاتها المأخوذة كمحكّ. ويرى أوبنهايم أنه لضمان الصدق في أسئلة الحقائق يمكن اللجوء إلى الصدق المحكي عن طريق ما يسمّى بالمراجعة المتقاطعة أو المتقابلة (Across – checks) حيث يكون هناك مصدر ثان مستقل للمعلومات وتتم المقابلة بين المعلومات المتحصّلة من الاستبانة وتلك المستمّدة من ذلك المصدر. ويمكن استخدام الوثائق والسجلاّت المختلفة كسجلاّت الخدمة العسكرية والدوائر الحكومية الأخرى، كما يمكن الاستعانة بالوالدين والمعلّم كمصدر للمعلومات، أي كمحك للصدق. ومن الباحثين من يقترح إجراء مقابلة مع عيّنة من المستجيبين للتحقّق من الصّدق ورفع مستواه في استخدام الاستبانة البريدية خاصة. وقد أشار أوبنهايم إلى أنه يمكن رفع مستوى الصدق بإجراء مقابلة شخصية مباشرة مع المستجيب وطرح مجموعة من الأسئلة عليه بسرعة قصوى ممّا يحفّزه على إعطاء معلومات حقيقية (Oppenheim,1966). وعموماً تعد مسألة الصدق هي المسألة المركزية ضمن طائفة واسعة من المسائل التي تثيرها عملية تطوير الاستبانة. ومن المفيد استخدام أكثر من طريقة واحدة في اختبار الصدق واللجوء إلى الصدق التجريبي بالإضافة إلى الصدق المنطقي.

الموثوقية والكفاية:

لا شك أن عملية التحقّق من صدق الاستبانة، على أهميتها، ليست دليلاً كافياً على صلاحية هذه الأداة، ولابد من التثبّت من موثوقيتها (ثباتها) بالإضافة إلى كفايتها لاستكمال عملية التحقّق من صلاحيتها. ويشير الثبات إلى اتساق البيانات والحصول على الإجابات نفسها ثانية في حال إعادة تطبيق هذه الأداة على حين أن الصدق يعبّر عمّا إذا كانت البنود (الأسئلة) تقيس فعلاً ما

وضعت لقياسه . ومن المعلوم أن أداة القياس قد تكون عالية في ثباتها وضعيفة في صدقها، فالأسئلة التي يجاب عنها بعبارات مقبولة اجتماعياً قد تحقّق ثباتاً عالياً مع أنّها ليست صادقة. إلاّ أن الصدق لا يمكن أن يصل إلى درجة معيّنة إذا لم تكن الأداة ثابتة بدرجة ما. وبعبارة أخرى لكي تكون الأداة على درجة عالية من الصدق لابد أن تكون عالية في ثباتها دون أن يدل ارتفاع مستوى الثبات بحد ذاته على صدق هذه الأداة. ومن هنا فإن دراسة الثبات (والتغيّر) في استجابات المبحوثين لا تنفصل عن دراسة مسألة الصدق، والواقع أن صدق الاستجابة (الاستجابات) التي يقدّمها المبحوثون تؤثّر تأثيراً بالغاً ومباشراً في ثبات هذه الأداة كما تؤثر بطبيعة الحال في صدقها، وبالتالي فإن دراسة الثبات والصدق تتطلّب مراجعة تلك الاستجابات. ومن المفيد الإشارة هنا إلى أن استجابة المبحوث قد لا تكون حقيقية أو «منزّهة» أو صادقة في كل الحالات فقد تتأثّر بالتفكير الرغبي وبالميل إلى إرضاء الباحث والرغبة في أن يكون شخصاً حسن الصّفات بنظر نفسه ونظر الآخرين، كما قد تتأثّر بخداع الذاكرة (سيفكينا، 1986). وثمّة أسباب عديدة تؤدّي إلى إضعاف صدق الاستجابة وتؤدّي بالتالي إلى إضعاف ثبات الاستبانة وصدقها منها أن المبحوث قد لا يعرف الإجابة عن السؤال فيلجأ إلى التخمين، أو أنه قد لا يفكّر في الأسئلة ولا يتأمّلها جيداً. هذا بالإضافة إلى أن المبحوث قد لا يفهم التعليمات جيداً، وقد يخشى قول الصدق أو يشعر أن السؤال شخصي جداً في طبيعته (جابر، 1987) . ويتم التثبّت من صدق الاستجابة (الاستجابات) بعدّة طرق منها المراجعة الداخلية للاستجابات (Internal Checks). ومن أشكال هذه المراجعة تقديم بنود وهمية أو زائفة كتقديم اسم أو برنامج إذاعي غير موجود، فإذا أقرّ المستجيب مثل هذه البنود فإن هذا يدل على لجوئه إلى التخمين ولا مبالاته. ومن أشكالها أيضاً إعادة طرح السؤال في موضع آخر من الاستبانة بعد تعديل صياغته أو وضعه في سياق آخر. فلو طرح السؤال التالي: (هل تذهب إلى المتحف؟) فيمكن دون إعادة السؤال نفسه، والتي غالباً ما تزعج المبحوث، وضع هذا السؤال في سياق آخر ليصبح بالشكل التالي: (هل تزور الأماكن التالية: السينما، حديقة الحيوان، المتحف، المعرض الفني؟). وفي بعض الحالات تتم المراجعة

الداخلية بطريقة منطقية فمثلاً إذا ادّعى شخص ما انه خاض الحرب فلابد من التأكّد انه كان في عمر معيّن وقت الحرب (Oppenheim, 1966). وغني عن البيان أن الطرائق المتّبعة في حساب الثّبات من مثل استعمال نماذج بديلة وإعادة إجراء الأداة ثم الربط بين النتائج يمكن استخدامها في التحقّق من ثبات الاستبانة. ومن الأجدى للباحث الاّ يقتصر على طريقة واحدة للتحقّق من الثبات فلكل من الطرائق المتّبعة في حساب الثّبات مزاياها الخاصة. وتدعم كل منها الأخرى وتكملها ممّا يسهم بالنتيجة في تأكيد الثّبات وتعزيزه وتقريره بأدلّة واضحة.

وبالإضافة إلى الثّبات والصدق لابد أن تحقّق الاستبانة شرط الكفاية كغيرها من الأدوات وأن يأخذ الباحث بالحسبان مسألة الاقتصاد في الوقت والجهد والنفقات ويعمل على إنقاصها إلى الحدود الدنيا شريطة الايلحق ضرراً بصدق الاستبانة وموثوقيتها. وعموماً تعدّ الاستبانة البريدية والاستبانة في شكلها المغلق أكثر كفاية من الاستبانة غير البريدية والأسئلة المفتوحة. ومن الضروري أن يتم اختيار الباحث لنوع الاستبانة وشكلها من منطلق مراعاة الشروط الثلاثة لصلاحيتها جميعاً مع تحقيق نوع من «التوازن» بين تلك الشروط ودون أن يكون الاهتمام بأحدها على حساب الآخر.

عيوب الاستبانة ومحاذيرها:

تفيد مراعاة الشروط الضرورية لتصميم الاستبانة والقيام بجملة الإجراءات اللازمة للتثبّت من صلاحيتها في تحاشي الكثير من مصادر الخطأ ولكنها لا تنفي كلية احتمالات الخطأ الذي يمكن أن يلحق بهذه الأداة الهامة من أدوات البحث.

وثمّة مصادر عديدة للخطأ تنعكس بصورة مباشرة أو غير مباشرة على الاستبانة ونتائجها وقد تؤدّي إلى إفسادها والتشكيك بها كأداة إذا لم يتنبّه

الباحث إليها ويعمل على تداركها أو التخفيف منها إلى الحدود القصوى. فإهمال المستجيبين أو سوء فهمهم لبنود الاستبانة، وارتفاع نسبة المستجيبين بما يتجاوز حدود التوقع، بالإضافة إلى أخطاء العيّنة والأخطاء التي يمكن أن تنجم عن عملية التفريغ والتحليل الإحصائي واستخلاص النتائج وغيرها من المؤثّرات يمكن أن تعمل منفردة أو مجتمعة وتقلّل من قيمة الإجراءات والخطوات المتّبعة في تصميم هذه الأداة والتحقّق من صلاحيتها، كما قد تؤدّي إلى التشكيك بمصداقيتها وقيمتها وتضعف وزنها كأداة بحث. ومن هذه الزاوية لابد أن تؤخذ مصادر الخطأ المشار إليها بالإضافة إلى أخطاء التصميم والانخداع ببساطة هذه الأداة وسهولتها الظاهرية على أنها من أهم الأسباب الكامنة وراء التشكيك بفاعلية هذه الأداة والإشارة إلى محدوديتها.

وتغطّي التحفّظات والانتقادات الشديدة الموجّهة لهذه الأداة جوانب عديدة تتّصل بالطبيعة الخاصة لها وصلاحيتها، والشروط الواقعية المحيطة بإجرائها والإفادة من نتائجها. ومن الباحثين من يرى أنه من الصعب الحصول على استجابات صادقة نظراً لأن استجابات المبحوثين تخضع لمؤثّرات سلبية عديدة وشديدة الفعالية . ويؤدّي ضعف مستوى الصدق في استجابات المبحوثين إلى عدم اتساق البيانات والتعارض بين الأقوال والأفعال وقد تكون بعض الاستجابات من نوع التعميمات الجامدة والتي قد تنطوي على شيء من الحذر والتحفّظ في قول الحقيقة و«مراعاة» السلطة. وقد أشار شوفيلد في هذا الصدد إلى أن «الكثير من الإجابات يحتمل أن تكون نمطية أو أن تحمل طابعاً استعراضياً خادعاً حتى عندما يبذل المبحوثون قصارى جهدهم للإجابة عن الأسئلة بعناية ويفكّرون فيها مليّاً». (Shofield, 1972, p.193). وتبعاً لذلك يصعب الوصول إلى أكثر من تعميمات أولية و«غير نهائية» من بيانات الاستبانة ، كما يصعب القول أنها تفضّل طريقة المقابلة في صدقها وثباتها مع الأخذ بالحسبان أن طريقة المقابلة تعاني من تدنٍّ واضح في مستوى الصدق والثبات.

(Shofield, 1972 ليرنر، 1983)

بالإضافة إلى ذلك يصعب عن طريق الاستبانة الكشف عن دوافع المبحوث ومشاعره وحالته النفسية. وليس من النادر «أن يحار الباحث عندما يجد أن المبحوثين يستخلصون معاني مختلفة كثيرة من الأسئلة التي اعتقد أنها واضحة ومحدّدة بصورة كافية» (Van Dalen, 1973, p.327).

كما أنه ليس من النادر في حال استخدام الأسئلة المغلقة الاّ يتمكّن الباحث من تقديم عدد من الإجابات البديلة تتيح للمبحوث ان يعبّر بدقّة وبصورة ملائمة بسبب من صعوبة التنبؤ باستجابات المبحوث (المبحوثين) المحتملة. ويؤدّي هذا الأمر إلى تقيّد المبحوث بطائفة محدّدة من الإجابات وقد يدفعه إلى إحداها. وقد تتلوّن الأسئلة بالتحيّزات الشخصية للباحث وتكون مشحونة باتجاه واحد، وقد تطرح في أوقات وظروف غير مواتية. ومن الصّعب على الباحث أن يضمن تعاون المبحوث لعدم وجود احتكاك شخصي مباشر بينهما ونظراً لأن المبحوث يتعامل مع كرّاسة صمّاء من الأسئلة وقليلاً ما يتعامل مع الباحث نفسه على حين انه يمكن أن يتم هذا التعاون والتجاوب في حالة الملاحظة أو المقابلة الشخصية المباشرة (جابر، 1987، يرماكوفا، 1979). ومن الانتقادات الموجّهة للاستبانة أيضاً أنها تصلح للدراسات التي تهتم بجمع بيانات عن المظاهر الخارجية للسلوك ولا تصلح للدراسات التي يحتاج فيها الباحث إلى بيانات تتّسم بالعمق والفهم لطبيعة الظاهرة (أو الظاهرات) المدروسة. (ماتيوشكين، 1981، كابانوفا، 1982) . ويستتبع هذا أن تقتصر الاستبانة على الموضوعات التي كوّن عنها المبحوث نظرة محدّدة وواضحة ويستطيع أن يعبّر عنها ببساطة (جابر، 1987). وغني عن البيان أن الاستبانة لا تصلح لصغار السن والأميين وأنها تتطلّب أكثر من مجرّد القراءة والكتابة.

التشغيل الأمثل للاستبانة وفاعليتها في البحث التربوي:

إلاّ أن التحفّظات والانتقادات الموجّهة للاستبانة والتي تلح على محدوديّتها وعيوبها الخاصة لا تنفي أهمية هذه الأداة ويقابلها العديد من نقاط القوة التي تبرز محاسنها ومزاياها الخاصة. فالاستبانة بوصفها أداة لجمع المعلومات يمكن استخدامها بصورة اقتصادية وفعّالة في التربية بمجالاتها العديدة والمتنوّعة كما يمكن عن طريقها الحصول على معلومات وبيانات يصعب الحصول عليها بالأدوات الأخرى. ومن مزايا هذه الأداة أنّها تأخذ شكلاً مقنّناً حيث تتم صياغة الأسئلة والتعليمات بطريقة واحدة للمبحوثين جميعاً بخلاف المقابلة الشخصية. وتضمن الصياغة المقنّنة للاستبانة والترتيب الثابت لأسئلتها والتعليمات المقنّنة لتسجيل الاستجابات قدراً كبيراً من الاطراد من موقف لآخر. ومع أن السؤال الواحد «قد تكون له معاني مختلفة عند الأشخاص المختلفين فإنه يمكن العمل على ضمان اطراد المعنى بالدراسة الاستطلاعية ومساعدة الأفراد على فهمه في أثناء تطبيقه» (جابر، 1987، ص 265). من جهة أخرى تقتضي تعليمات الاستبانة عادة عدم كتابة الاسم ممّا يفسح المجال لتوجيه عدد من الأسئلة التي قد يصعب طرحها بوسيلة أخرى، ويعطي المستجيبين فرصة التعبير عن آرائهم ومواقفهم التي قد يخشون عدم موافقة الآخرين عليها (جابر، 1987، التير، بلا تاريخ).

وتتفوّق الاستبانة على طريقة المقابلة من حيث أنّها أداة اقتصادية في الوقت والجهد والنفقات ويمكن استخدامها في المواقف الجمعية وتطبيقها على أعداد كبيرة من الناس في وقت واحد. ثم أنّها تحتاج لمهارة أقل من المقابلة ويمكن إرسالها بالبريد إلى أفراد ينتشرون في رقعة جغرافية واسعة. ويرى جرونلند أن الاستبانة أداة مجدية في عملية التوجيه والإرشاد «حيث أن رغبة الفرد في أن يفهم نفسه ويتّخذ قرارات مستقبلية حكيمة تشجّعه على أن يعرف نفسه كما هي عليه في الواقع وليس كما يرغب هو أن تكون أو كما يرغب أن يراه الآخرون» .(Gronlund, 1971, p.19)

واخيراً فإنّ المفاضلة بين أدوات البحث واختيار الباحث لأداة معيّنة (أو أكثر) دون غيرها يجب أن يتم في ضوء طبيعة البحث وأغراضه الخاصة ومستلزماته. وغني عن البيان أنّ لكل أداة من أدوات البحث التربوي مزاياها وعيوبها. إلاّ أن هذه المزايا وما يقابلها من العيوب يجعل من الصعب الحكم على افضلية أداة معينة على سواها بصورة قاطعة وفي كل الحالات نظراً لأن الأداة، أي أداة، تخضع لطائفة واسعة من العوامل والمتغيرات. وهذه العوامل أو المتغيّرات يصعب عزلها أو تثبيتها وبالتالي يصعب فصلها عن الظاهرة المدروسة في واقعها الخاص ممّا ينعكس بتأثيرات مختلفة على الأداة ذاتها من حالة لأخرى وقد يعزّز مزاياها (أو عيوبها) حيناً وقد يضعفها حيناً آخر. واختيار الباحث للاستبانة كأداة للبحث، أو كواحدة من أدواته، لابد أن يتم من هذا المنطلق وفي ضوء الدراسة المعمّقة لهذه الأداة ومزاياها وفاعليتها في الوضع الخاص وبعد التأكّد من افضليتها وتفوّقها على تلك التي لن تحظى بالاختيار. من جهة أخرى فإن السؤال الأساسي حول جدوى الأداة، أي أداة وفاعليتها لا ينفصل عن درجة جودة هذه الأداة ومواصفاتها الفنية الضرورية المعبّرة عن صلاحيتها. ومن الضروري العمل على تأمين صلاحية الأداة منذ البدء بعملية تصميمها وفي أثناء سير هذه العملية بكافة مراحلها وخطواتها لكي يتمكّن الباحث من «تشغيل» هذه الأداة و«استثمارها» والإفادة من مزاياها إلى الحدود القصوى. وتبعاً لذلك يكون توفير الشروط الضرورية لتأمين صلاحية الاستبانة جزءاً لا يتجزّأ من عملية تصميمها، أو أن هاتين العمليتين تكونان بمثابة جانبين لعملية واحدة متكاملة تتّجه إلى تحقيق الإفادة القصوى من مزايا هذه الأداة والتقليل من عيوبها ما أمكن ذلك. وعلى هذا لابد من توفير مستلزمات صلاحية الاستبانة منذ المرحلة الأولى لعملية تصميمها واختبار هذه الصلاحية (حسب الحاجة) ورفعها إلى الحدود القصوى من خلال إخضاعها للتطبيق التجريبي بالإضافة إلى التحليل المنطقي.

وأخيراً فإن الاستبانة بوصفها أداة للبحث والقياس التربوي وطريقة في الحصول على المعلومات والبيانات قد تكون أداة وحيدة وتفي بأغراض البحث

كاملة في حالات، وقد تكون أداة «متممة» في حالات أخرى دون أن ينتقص هذا من قيمتها أو يضعف من وزنها ومكانتها.

ويمكن أن تؤدّي الاستبانة دورها على النحو الأمثل إذا تضافرت مع بقية الطرائق والأدوات أو بعضها كالملاحظة بأنواعها والاختبارات والمقابلة الشخصية. وسواء أكانت الاستبانة أداة وحيدة أو «متممة» فإن المسألة الملحّة المطروحة هي التحقّق من صلاحيتها و«تشغيلها» بالصورة الملائمة وبما يكفل تحقيق كافة الفوائد التي يمكن أن تقدّمها.

خلاصة:

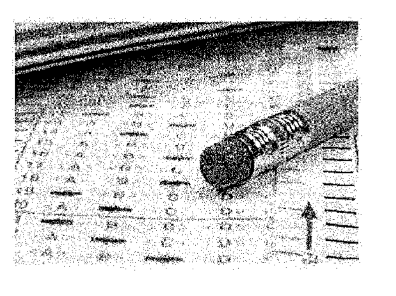
تركز الاهتمام في هذا الفصل على إلقاء الضوء على الاستبانة وموقعها في البحث والقياس النفسي والتربوي وفاعليتها كأداة من أدواته، كما اهتم بمعالجة طائفة واسعة من المسائل المتّصلة بشكلها ونوعها وعملية تصميمها وتوفير مستلزمات صلاحيتها والتثبّت من هذه الصلاحية.

وثمّة تفاوت واسع في وجهات نظر الباحثين بصدد الاستبانة وقيمتها كأداة بحث وقياس. والواقع أن الكثير من البحوث التربوية، بما في ذلك البحوث المعدّة لنيل أعلى الدرجات العلمية، تعاني من عدم مراعاة الشروط الضرورية لتصميم الاستبانة، ومن غياب مؤشّرات صلاحيتها بصورة كلّية أو جزئية. وينطلق الفصل الحالي من أن لكل أداة بحث وقياس ميزاتها وأفضليتها تبعاً لأغراض البحث، وأن الاستبانة إذا توافرت فيها شروط الصلاحية فإنّها يمكن أن تكون أداة بحث فعالة وأن تغطّي مجالات واسعة يصعب أن تطالها بقية الأدوات. بالإضافة إلى ذلك فإن التحليل المنطقي لبنود الاستبانة ليس عملاً متمّماً أو لاحقاً بها بل هو عمل مستمر ومواكب لعملية تصميمها. ومن هذه الزاوية لا يجوز أن نعطي الأولوية للعمل التجريبي والتحليل الإحصائي.

والنتيجة المستخلصة من هذه الدراسة هي أن السؤال حول جدوى هذه الأداة وفاعليتها لا ينفصل عن مسألة صلاحيتها. وتبعاً لذلك يكون توفير الشروط الضرورية لتأمين صلاحية الاستبانة جزءاً لا يتجزأ من عملية تصميمها. ولابد من اختبار هذه الصلاحية ورفعها إلى الحدود القصوى من خلال إخضاع الاستبانة للتحليل المنطقي من قبل خبراء ومختصين يتعاملون معها بصورة جدية ، وعدم الاقتصار على "التحكيم الشكلي " كما يحدث في كثير من الاحيان. هذا بالإضافة إلى التطبيق التجريبي والتحليل الإحصائي بطبيعة الحال.

الفصل الرابع

بناء اختبارات التحصيل



الفصل الرابع
بناء اختبارات التحصيل

مقدمة:

تمثّل اختبارات التحصيل الأداة الوحيدة للتقويم في الأنظمة التربوية التي توصف عادة بالتقليدية. وقد وجّهت للاختبارات التحصيلية بصورتها التقليدية خاصة انتقادات شديدة، واتجه المربّون نحو تطوير أشكال جديدة لها بهدف التخفيف من عيوبها، كما تعزّز الاتجاه الداعي إلى عدم الاقتصار عليها في عملية التقويم وضرورة استخدام طرائق ووسائل جديدة للتقويم تلبي الأهداف الأساسية للتربية الحديثة. وعلى الرغم من ذلك كله فإن اختبارات التحصيل ما زالت تحتل المكانة الأولى في عملية التقويم. وتكتسب مسألة تحسين هذه الاختبارات وتحديثها شكلاً ومضموناً ورفع فاعليتها أهمية خاصة في ظروف الثورة العلمية والتقنية، ومع التوسّع الهائل في حجم المعارف العلمية الذي يشهده عصرنا، حيث بات من الضروري إعداد الإنسان الجديد القادر على استيعاب منجزات العلم والتقنية المتطورة باستمرار.

وسنحاول في هذا الفصل أن نلقي الضوء على الاختبارات التحصيلية ودراسة المسائل العديدة المتصلة بعملية بنائها وتوفير مستلزمات صلاحيتها، بهدف إحاطة القارئ بها من جهة، وتمكين المعلّم خاصة من جهة أخرى، من تحقيق الفائدة القصوى من هذه الاختبارات واستعمالها على النحو الملائم وبما يؤدي إلى رفع فعالية عملية التعلّم والتعليم.

الانتقادات الموجّهة للامتحانات التقليدية بوصفها مقاييس تحصيلية:

من المؤكد أن مشكلة تسلّط الامتحانات بصورتها التقليدية هي إحدى المشكلات الكبرى التي تواجهها الأنظمة التربوية المختلفة في الكثير من بلدان العالم منذ زمن بعيد وحتى يومنا هذا. فالامتحانات التقليدية تقتصر على قياس التحصيل الدراسي بل وتؤدي إلى تضييق مفهوم التحصيل الدراسي، بحيث ينصب على مقدار ما يسترجعه المتعلّم من معلومات دون إحداث تغيرات سلوكية في اتجاهات المتعلّم وقيمه وشخصيته ككل. وهذا ما يتعارض مع الأهداف الأساسية للتربية الحديثة التي تتركّز في التطوير المتسق والمتكامل لشخصية المتعلّم، وتنمية هذه الشخصية من كل النواحي بدءاً بالناحية الجسمية ومروراً بالنواحي العقلية والمعرفية وانتهاء بالنواحي الوجدانية والخلقية والاجتماعية وغيرها. وقد وجّهت للامتحانات التقليدية انتقادات لاذعة وظهرت اتجاهات عديدة تدعو إلى التخفيف من وطأتها وعيوبها. وما من شك في أن مسألة الامتحانات هي مسألة كبيرة بحد ذاتها، ومتعدّدة الجوانب والأبعاد. وليس من قبيل المبالغة القول: إن هذه المسألة تعدّ المسألة الأولى والأكثر إلحاحاً و «إشكالية» في واقعنا التربوي الراهن.

والانتقادات التي توجّه إلى الامتحانات التقليدية، والإشكالات التي تثيرها عديدة ومتنوّعة، وقد يكون من الصعب حصرها جميعاً وتقديم صورة وافية أو شاملة لها. والمقصود بالامتحانات التقليدية تلك الامتحانات التي تأخذ شكل الاختبارات المقالية أو الإنشائية كما تأخذ في بعض الأحيان شكل الاختبارات الشفهية. وهذه الأخيرة أقل شيوعاً في الوطن العربي منها في بعض البلدان (كبلدان شرقي أوروبا مثلاً). كما يقصد بالامتحانات التقليدية الامتحانات أو الاختبارات التي يجريها المعلّم لأغراض الاستعمال الصفي، والامتحانات النهائية بما في ذلك الامتحانات الرسمية وامتحانات الشهادات. ويمكن إجمال الانتقادات الأساسية للامتحانات التقليدية فيما يلي:

1. إن الامتحانات التقليدية كثيراً ما تقتصر على قياس المستوى الأدنى من مستويات التعلّم وهو مستوى المعرفة الذي يعتمد على الحفظ والاسترجاع بصورة أساسية، وتهمل المستويات الأعلى للتعلّم كالفهم والتطبيق والتحليل والتركيب وغيرها، أو تعطيها دوراً ثانوياً للغاية. ولهذا الأمر انعكاساته الخطيرة على عملية التعلّم والتعليم، وعلى التكوين المعرفي والفكري للطالب. فالأسئلة التقليدية إذ تركّز على قياس ما حفظه الطالب من معلومات تهمل الجانب الأهم في عملية التعلّم والتعليم وهو تنمية قدرة الطالب على استخدام أو تطبيق هذه المعلومات، وربطها بالحياة العملية، كما تهمل مسألة التقدّم الفكري للطالب. وهي بذلك تعد محدودة في قيمتها نظراً لأنها تركّز على مظاهر سطحية وغير هامة في التحصيل.
2. تعجز الأسئلة التقليدية عن تقديم عيّنة ممثّلة لمحتوى المنهاج الدراسي، وهذا ما يقلّل من صدق تلك الاختبارات (صدق المحتوى) ويزيد من أثر عامل المصادفة في النجاح والرسوب أو ما يسمى "يانصيب الامتحانات".
3. عدم وجود خطة محدّدة وواضحة تربط عملية القياس والتقويم بالأهداف التعليمية ممّا يؤدي إلى خلق هوة عميقة بينهما، ويضعف بدوره من صدق تلك الاختبارات من جهة قدرتها على تتبّع مدى تقدّم المفحوص نحو الأهداف التعليمية المرسومة. (السيد ومخائيل، 1989).
4. تقتصر تلك الاختبارات إلى الموضوعية والثبات وتبرز فيها بشكل واضح مشكلة التقديرات والعوامل الشخصية في التصحيح والتي يعبّر عنها بالاختلاف بين مصحّح وآخر في إعطاء الدرجات (العلامات)، أو حتى عند المصحّح نفسه من وقت لآخر. وهذا ما يدعو إلى التشكيك بقيمة تلك الاختبارات بوصفها أدوات غير دقيقة وغير حسّاسة بصورة كافية للفروق الحقيقية في التحصيل.
5. تهمل الاختبارات التحصيلية بصورتها التقليدية الأغراض التشخيصية للقياس أو تعطيها وزناً ضئيلاً، وكثيراً ما تعجز عن تقديم صورة

تشخيصية عن مَواطن الضعف والقوة في أداء المفحوص. وتسعى هذه الاختبارات إلى تلبية الأغراض الإدارية للقياس والتي تتمحور في إعطاء الدرجات والتقديرات، وتمثّل بذلك الأساس الوحيد في اتخاذ القرارات الإدارية المتصلة بالمستقبل الدراسي والمهني للطالب مع أنها تعاني أساساً من ضعف مستوى الصدق والثبات، ومن عيوب التقديرات الذاتية التي تجعلها مقاييس غير دقيقة و«غير موضوعية» بصورة كافية لقدرات الطالب الحقيقية كما أسلفنا.

وبالإضافة إلى الانتقادات التي توجّه إلى الامتحانات التقليدية بوصفها أدوات لقياس التحصيل، تثير هذه الامتحانات مسائل وإشكالات عديدة منها، أنها تجعل عملية التعلّم موجّهة بفكرة الامتحان وخاضعة لها بحيث يتركّز الهدف الأساسي للمتعلّم على النجاح في الامتحان، والحصول على أعلى الدرجات بغض النظر عمّا تعلّمه واتقنه فعلاً من المادة الدرسية. وكثيراً ما تُعرّض هذه الامتحانات الطلاب المفحوصين لخبرات الفشل والإحباط والقلق ممّا يضعف ثقتهم بأنفسهم، ويهدّد مستقبلهم الدراسي والمهني. هذا بالإضافة إلى أنها تثير الروح التنافسية وقد تعزّز الميول الفردية والأنانية بين الطلاب، كما قد تؤدي إلى إشاعة الفوضى وانتهاك النظام المدرسي.

ومهما يكن من أمر فإن الامتحانات التقليدية تعاني من نواقص وعيوب كثيرة. وقد يكون من المفيد بدلاً من التركيز على تلك النواقص والعيوب، التركيز على كيفية تحسينها، والتخفيف من مساوئها ومضارها، والاستعانة بوسائل وأدوات تقويمية أخرى تضاف إليها. فالامتحانات والاختبارات بأنواعها تمثّل جانباً أساسياً وبالغ الأهمية في العملية التعلّمية – التعليمية أو جزءاً لا يتجزأ منها، ويستحيل بالتالي الاستغناء عنها أو إقصاؤها من هذه العملية. ولكن من الضروري إقصاء تلك العيوب والنواقص التي قد تعود في معظمها إلى سوء استعمالنا لها، وعجزنا عن توظيفها بالاتجاه المطلوب. وما من شك في أن تحسين

الامتحانات يجب أن يبدأ بالاختبارات الصفية التي يعدّها المعلّم والتي تشغل الحيز الأكبر والأهم في عملية القياس التقويم والتربوي.

بناء اختبارات المعلّم الصفية:

المقصود باختبارات المعلّم الصفية تلك الاختبارات التي يعدّها المعلّم عادة لأغراض الاستعمال الصفي، وتطبّق في أثناء سير عملية التعلّم والتعليم (خلال الفصل الدراسي أو العام الدراسي) وبانتهاء هذه العملية (نهاية الفصل الدراسي أو نهاية العام الدراسي). وتنطوي هذه الاختبارات على أهمية قصوى نظراً لأنها تشغل الحيّز الأكبر في عملية تقويم الطالب، وتكاد تكون الأداة الوحيدة للتقويم في الكثير من أنظمة الامتحانات والتقويم السائدة حتى يومنا هذا. ومن المعلوم أن الاختبارات الصفية تعاني من ضعف التخطيط وربما غيابه كلياً، وقليلاً ما تخضع بنود هذه الاختبارات للتحليل المنطقي، ونادراً ما تخضع للتحليل الإحصائي أو لا تخضع له على الإطلاق. ولعل السبب في ذلك هو عدم إلمام المعلّمين بصورة كافية بنظرية القياس المعاصرة وممارساته المتقدّمة ، واعتقاد الكثير منهم أن معرفتهم الجيدة للمادة العلمية يمكن أن تؤهلهم للقياس بصورة تلقائية. هذا بالإضافة إلى أن عملية قياس السلوك الإنساني تعد عملية معقّدة بحد ذاتها. وما من شك في أنه من الخطأ الاستهانة بعملية إعداد الاختبار الصفي والنظر إليها وكأنّها مهمة إضافية أو ثانوية بحجّة أن التدريس، والذي يتمثّل بإعطاء المعلومات المقرّرة وشرحها، يعد بمثابة المهمّة الأساسية للمعلّم. فالتدريس والتقويم جانبان لعملية واحدة متكاملة، والتقويم الفعّال هو عنصر متمّم للتعليم الفعّال ودليل عليه. وترتكز عملية القياس والتقويم الصفي على مجموعة من الأسس والمبادئ التي لابد من مراعاتها عند بناء الاختبار الصفي. وانطلاقاً من تلك الأسس والمبادئ تسير عملية بناء الاختبار وفق خطوات محدّدة من أهمّها : التخطيط للاختبار، وكتابة البنود الاختبارية، وإخضاعها للتحليل المنطقي والإحصائي. ومن المؤكد أن الوقوف عند هذه الأسس والمبادئ وكذلك المراحل والخطوات الأساسية في عملية بناء الاختبار ودراستها بصورة متأنية يمكن

أن تحقّق فائدة قصوى للمعلّم وقد تكون بمثابة دليل عمل له في نشاطه التقويمي الصفي.

أسس بناء الاختبار الصفي:

من المبادئ والأسس الهامة في القياس الصفي، والتي يمكن أن يؤدي العمل بموجبها إلى تحسين الاختبارات الصفية بصورة ملحوظة، يمكن أن نذكر المبادئ والأسس التالية التي أكدها الكثير من علماء القياس من مثل جرونلند ومهرنز وأنستازي وغيرهم.

1. إن عملية بناء الاختبار يجب أن تأخذ بالحسبان استعماله والغرض الخاص منه. فإذا كان الغرض من الاختبار هو تقويم تقدّم التلاميذ نحو الأهداف التعليمية المرسومة فلابد أن يوضع الاختبار بحيث يلبّي هذا الغرض، ويكشف عن إنجاز التلاميذ ومدى تحقيقهم لتلك الأهداف. وإذا كان الغرض منه هو الكشف عن صعوبات التعلّم فلا بد أن يتمتّع بقدرة تشخيصية عالية، ويكون محدوداً في مجاله، كما لابد أن تميل بنوده إلى السهولة وتوضع في ضوء مصادر الخطأ الأكثر شيوعاً لدى التلاميذ. وإذا كان الغرض من الاختبار هو قياس التمكن أو الإتقان لوحدة دراسية ما فلا بد أن يختص بهذه الوحدة دون غيرها ويرتكز على محك للتمكّن أو الإتقان يتم في ضوئه تقرير ما إذا كان الطالب قد أتقن هذه الوحدة أم لا.
2. إن اختيار شكل البنود (أو الأسئلة) يتحدّد بطبيعة الأهداف ونواتج التعلّم المقيسة. ومن المعلوم أن كل نوع أو شكل من أشكال البنود الاختبارية قد يكون أكثر ملاءمة لقياس نواتج تعلّمية معيّنة، وأقل ملاءمة لقياس نواتج أخرى. فالاختبار المقالي أكثر فعالية من الاختبار الموضوعي في قياس القدرة على الربط وتنظيم الأفكار. والاختبار الموضوعي من نوع التكميل أو تعبئة الفراغ أكثر ملاءمة من سواه لقياس القدرة على استدعاء الحقائق الخاصة. واختبار الاختيار المتعدّد يصلح لقياس الكثير من نواتج التعلّم

ومستوياته الدنيا منها والعليا. وما من شك في أن الاستخدام الفعّال لكل نوع أو شكل يتطلب مهارة خاصة من واضعه. فالاختبار المقالي قد يقتصر على قياس القدرة على الاسترجاع، ولا يكشف عن المستويات العليا للتعلّم، وكذلك الأمر فيما يتصل باختبار الاختيار المتعدّد الذي قد يقيس المعرفة والحفظ فقط إذا لم تعط العناية الكافية لعملية إعداده.

3. يجب وضع البنود الاختبارية بحيث تكون بمثابة عينة ممثّلة لمحتوى المادة ونواتج التعلّم الخاصة المقيسة. فلكل مجال من مجالات المحتوى، ولكل نتاج تعلّمي خاص لابد من اختيار عيّنة من البنود تؤخذ الإجابة عنها بمثابة دليل على التحصيل في ذلك المجال. ويفترض عادة أن إجابة التلميذ عن تلك البنود أو الأسئلة تمثّل إجاباته المحتملة عن كل البنود أو الأسئلة المحتملة. فنحن نتوقّع من التلميذ معرفة آلاف الحقائق الخاصة ونختبره بعدد محدود منها، ونتوقّع أن يطوّر مهاراته في حل طائفة واسعة من المشكلات ونختبره بإحداها أو بعدد محدود منها. ويعد جدول المواصفات وسيلة من أهم الوسائل في الحصول على عيّنة ممثّلة للمحتوى الدراسي ونواتج التعلّم المراد قياسها. وبطبيعة الحال فإنه كلّما ازداد عدد البنود اصبحت العينة أكثر تمثيلاً والنتائج أكثر ثباتاً. ويرى ن.جرونلند أنه لضمان الصدق يستحسن استخدام الأشكال المختلفة من البنود الاختبارية بما فيها أسئلة المقال، ولا يجوز التضحية بالصدق من أجل رفع مستوى الثبات عن طريق زيادة عدد البنود. (Gronlund, 1971, p.139).

وهذا يعني أن زيادة عدد البنود ترفع الثبات حتماً، ولكنها لا تضمن الصدق وبخاصة إذا كان الغرض الخاص بالاختبار الصفي هو عدم الاقتصار عن نواتج التعلّم البسيطة، والتصدي للنواتج المعقّدة أو العليا بالإضافة إلى النواتج البسيطة.

4. يجب أن تكون البنود الاختبارية من مستوى صعوبة ملائم. ومن المعلوم أن المستوى الملائم للصعوبة في اختبارات الإنجاز التي تنتمي عادة إلى فئة الاختبارات المعيارية المرجع هو 50٪. والبنود التي يصل معامل سهولتها إلى 0.50 تتمتّع بقدرة تمييزية عالية، وتحقّق مستوى عالياً من الثبات. إلاّ أن هذا لا يعني أنه يمكن استعمال البنود الغامضة أو تلك التي تتّصل بحقائق غير هامة من أجل رفع مستوى الصعوبة. فاستعمال مثل هذه البنود يؤدي إلى إضعاف الصدق من جهة، ولا يعد أمراً مستحباً من الناحية التعليمية من جهة أخرى نظراً لأن الطلاب سيلحّون في هذه الحالة على النقاط والجوانب غير الهامة في المادة ويهملون الهامة. والطريق الملائمة لرفع مستوى الصعوبة هو التوجّه نحو الأهداف والنواتج العليا كتطبيق المبادئ والمفاهيم وتحليل البيانات وغيرها. ولا تعد الصعوبة أمراً هاماً في اختبارات الإتقان والاختبارات التشخيصية حيث لا يكون الغرض الأساسي للاختبار هو إعطاء الدرجة أو العلامة الكلية للمفحوص وتعيين مركزه النسبي في المجموعة.

5. يجب وضع البنود الاختبارية بحيث لا تؤثر العوامل الجانبية أو الداخلية في أداء المفحوص، وتمنعه عن إظهار مستواه الحقيقي في التحصيل. «فالفروق في القدرة القرائية والمهارات الحسابية ومهارات الاتصال وما شابه، يجب ألا تؤثر في استجابات التلاميذ ما لم تكن هذه النواتج ذاتها موضوعاً للقياس» (Gronlund, 1971, p.141). ومن هذه الزاوية يجب الابتعاد عن الجمل الطويلة المعقّدة، والمفردات الصعبة، والكلمات التي تحمل أكثر من معنى، كما لابد من تجنّب الإيحاءات والتلميحات وكل ما يمنع الاختبار عن العمل بالاتجاه المطلوب، والكشف عن الفروق الحقيقية في أداء المفحوصين.

6. يجب وضع الاختبار بحيث يسهم بتحسين العملية التعلّمية – التعليمية. فالغرض النهائي للقياس الصفي هو تحسين تعلّم التلميذ. والاختبار الصفي الجيد يجب أن يرفع نوعية التعلّم وكميته، ويساعد المعلّم في تحسين أدائه التعليمي. وما من شك في انه إذا كان الاختبار عيّنة ممثّلة

للمحتوى والنواتج السلوكية المرغوبة فسيكون له أثره الإيجابي في تركيز اهتمام التلميذ بمجالات المحتوى، والنواتج المرغوبة كافة، وعدم الاقتصار على تذكّر الحقائق والمعلومات.

التخطيط للاختبار:

تعاني الاختبارات الصفية من ضعف التخطيط ، أو انعدامه كلياً كما أسلفنا، ممّا يؤدي إلى التشديد على بعض مجالات المحتوى الدراسي على حساب الأخرى، كما تؤدي إلى التشديد على البنود التي تتطلّب الاسترجاع البسيط للحقائق نظراً لسهولة إعدادها . وتتطلّب عملية إعداد الاختبار الصفي الجيد وضع خطة مسبقة وشاملة يتحدّد فيها الغرض من الاختبار أو الوظيفة (أو الوظائف) التي نتوخّى منه أن يؤديها، كما تتحدّد فيها موضوعات أو عناصر المحتوى، والأهداف التعليمية التي يسعى المعلّم إلى تحقيقها في المجال الذي يغطيه الاختبار. هذا بالإضافة إلى تحديد أشكال أو أنواع البنود الاختبارية التي تعدّ أكثر فاعلية في الكشف عن مدى تحقّق تلك الأهداف. ويقترح مهرنز مجموعة الأسئلة التالية التي يتعيّن على المعلّم، وكذلك المهني المختص ببناء الاختبارات، أن يطرحها على نفسه ويحاول الإجابة عنها في مرحلة التخطيط لبناء الاختبار. وقد نظّمت هذه الأسئلة بحيث تأخذ شكل قائمة رصد وذلك على النحو التالي:

الجدول رقم (3): قائمة رصد لمرحلة التخطيط في إعداد الاختبارات الصفية:

1)	ما الغرض من الاختبار؟ ولماذا أعدّه؟
2)	ما المهارات والمعارف والاتجاهات.. الخ التي أريد قياسها ؟
3)	هل تحدّدت أهدافي التعليمية بوضوح وبصيغ سلوكية؟
4)	هل اعددت جدول مواصفات؟
5)	ما نوع الاختبار (شكله) الذي أريد أن أستعمله ؟ ولماذا ؟

(6	كم سيكون طول الاختبار؟
(7	كم ستكون صعوبة الاختبار؟
(8	ما هو المستوى «التمييزي» لبنود الاختبار؟
(9	كيف سارتّب الأنواع (الأشكال) المختلفة من البنود؟
(10	كيف سارتّب البنود ضمن كل نوع (شكل)؟
(11	ماذا عليّ أن أفعل لكي أعد طلابي لأداء الاختبار؟
(12	كيف سيجيب الطلاب على الفقرات الموضوعية؟ على ورقة إجابة منفصلة أم على كراسة الاختبار؟
(13	كيف ستصحح البنود الموضوعية؟ باليد أم بالآلة؟
(14	كيف ساصحّح السؤال (الأسئلة) المقالي؟ بالطريقة الكلية أم التحليلية؟
(15	هل ستعطى تعليمات بالسماح بالتخمين في البنود الموضوعية؟ وهل ستطبّق معادلة التصحيح من أثر التخمين؟
(16	كيف ستتم جدولة العلامات؟
(17	كيف سيتم تعيين العلامات (الدرجات أو مستوى الأهلية)؟
(18	كيف سيتم تسجيل (تقرير) نتائج الاختبار؟

(Mehrens, 1973, p.176)

ويمكن النظر إلى الأسئلة السابقة بوصفها إطاراً نظرياً عاماً لعملية بناء الاختبار برمتها. وسيكون من المفيد الآن الوقوف عند الأسئلة الخاصة بمرحلة التخطيط السابقة لمرحلة وضع البنود وتحليلها، ودراستها بشيء من التفصيل.

تحديد الغرض من الاختبار:

فالاختبارات الصفية يمكن أن تلبّي العديد من الأغراض منها:

1. الحكم على إتقان معرفة أو مهارات أساسية معيّنة.
2. ترتيب الطلاب حسب تحصيلهم للأهداف التربوية.
3. تشخيص صعوبات التعلّم.
4. تقويم الطريقة التعليمية.
5. تحديد فعالية المنهاج.

وبالطبع فإن الاختبار الواحد يمكن أن يحقّق أغراضاً عديدة دفعة واحدة، ولكن لابد من تحديد هذه الأغراض بصورة مسبقة. فإذا رغب المعلّم مثلاً في التأكّد من إتقان تلاميذه لمعارف أو مهارات أساسية معيّنة، فإن الخصائص «السيكومترية» للبنود ستختلف عمّا لو كان يرغب في ترتيب طلابه حسب درجة إنجازهم وتحصيلهم أو تحقيقهم لتلك الأهداف. ففي الحالة الأولى سيهتم المعلّم بإعداد بنود سهلة نسبياً، وسيقتصر الاختبار على عينة محدودة أو ضيّقة للمحتوى. وفي الحالة الثانية سيهتم المعلّم بأن تكون أكثرية بنوده من مستوى متوسط من حيث الصعوبة، وأن يمثّل اختباره مدى واسعاً من محتوى المادة الدرسية. وفي كل الأحوال فإن القرار الخاص بالغرض من الاختبار يجب اتخاذه قبل كتابة البنود. وقد يكون من المفيد الإشارة هنا إلى أن الاختبار الواحد قد يحقّق أغراضاً كثيرة ولكن ليس بدرجة واحدة من الفاعلية. ومن هنا لابد من تحديد الغرض الرئيس مسبقاً لإتاحة الفرصة أمام المعلّم وطلابه لتحقيق الفائدة القصوى من الاختبار. (Mehrens, 1973)

تحديد الأهداف التعليمية:

إن الخطوة الأصعب والاهم في مرحلة التخطيط لعملية بناء الاختبار هي تحديد الأهداف التعليمية. فدون أهداف واضحة ومحدّدة لا يدري المعلّم ماذا يقيس، وقد يلجأ إلى الطريق الأسهل ويقتصر على أسئلة الاستدعاء البسيط مهملاً بذلك المستويات الأعلى للتعلّم. وما من شك في أن الأهداف التي تشكو من العمومية والغموض يصعب الاسترشاد بها في عملية التقويم. ومن هنا لابد من صياغة الأهداف بعبارات واضحة ومحدّدة والتعبير عنها بصورة نواتج سلوكية قابلة للملاحظة والقياس.

ويؤكد علماء القياس من أمثال جرونلند ومهرنز أن إخضاع الأهداف التعليمية للقياس يتطلّب السير بالخطوتين التاليتين:

1. صياغة الأهداف التعليمية العامة بلغة نواتج التعلّم المتوقّعة.
2. وضع قائمة بعيّنة من نماذج السلوك الخاصة التي يجب على الطالب أن يظهرها عندما ينجز الهدف.

وتتطلّب الخطوة الأولى التعبير عن الأهداف بصيغة نواتج أو «محصلات» عامة وغير سلوكية على الأغلب، على حين أن الخطوة الثانية تتطلّب الصياغة السلوكية للأهداف، وذلك من خلال وضع قائمة لعيّنة من نماذج أو أنواع السلوك التي تدل على تحقّق كل هدف (Mehrens,1973)

ففي الخطوة الأولى يمكن صياغة الهدف العام التالي مثلاً: «أن يفهم الطالب المبادئ العلمية»، وفي الخطوة الثانية لابد من وضع قائمة لنواتج التعلّم الخاصة التي يمكن أن تؤخذ دليلاً على تحقّق هذا الهدف. وهذه القائمة هي مجرّد عيّنة من النواتج التعلّمية التي تظهر فهم المبادئ العلمية ولا تشمل سائر النواتج بطبيعة الحال. ويتّضح ذلك على النحو التالي:

الهدف العام: أن يفهم التلميذ المبادئ العلمية (الخطوة الأولى).

النواتج الخاصة: (الخطوة الثانية)

1. أن يصف المبادئ العلمية بعباراته الخاصة.
2. أن يعطي أمثلة عنها.
3. أن يصوغ فرضيات ترتكز على هذه المبادئ.
4. أن يصنّف الفروق بين اثنين منها.
5. أن يشرح الصلات القائمة بين اثنين منها.

ومن الواضح أنه يمكن من حيث المبدأ وضع قائمة لكل نواتج التعلّم الخاصة، كما يحدث عادة في نطاق التعليم المبرمج أو في مجال تدريبي خاص، إلاّ أن هذا الأمر ليس ضرورياً في القياس الصفي ويتطلّب جهداً كبيراً ووقتاً طويلاً جداً. ومن هنا يمكن الاقتصار على عيّنة من النواتج شريطة أن تكون ممثّلة وصادقة. وهذه العيّنة لابد من ربطها بعناصر أو مجالات المحتوى الدراسي عن طريق جدول المواصفات.

تحديد المحتوى:

إن تحديد المحتوى الذي سيشمله الاختبار هو أمر هام للغاية وبخاصة حين يكون الغرض من الاختبار هو تقويم ما تعلّمه التلميذ في المادة الدرسية أو في جزء منها. ويرى ثورندايك وهاجن «أن للمحتوى أهمية خاصة لأنه الوسيط الذي تتحقّق من خلاله الأهداف كعمليات» (ثورندايك وهاجن 1986، ص 193). وتتطلّب عملية تحديد المحتوى بيان العناصر والمجالات الرئيسة والفرعية لهذا المحتوى مع تحديد الأوزان النسبية لكل منها، وذلك على ضوء أهميته، والزمن المخصّص لتدريسه لكي يصار إلى تخصيص عدد من البنود لكل مجال رئيس أو فرعي استناداً إلى ذلك.

جدول المواصفات:

بعد تحديد الأهداف التعليمية وعناصر المحتوى الذي سيشمله الاختبار لابد من تنظيمها جميعاً في جدول مواصفات الاختبار. والغرض من هذا الجدول هو ربط الأهداف التعليمية مباشرة بعناصر المحتوى الدراسي وتحديد عدد البنود الخاصة بكل هدف مرتبطاً بكل عنصر من عناصر المحتوى. ويفيد جدول المواصفات في إعطاء كل هدف الوزن الذي يستحقّه بنظر واضع الاختبار، ويعد تبعاً لذلك أداة فعّالة في تأسيس صدق محتوى الاختبار من خلال إلزام واضع الاختبار على توزيع أسئلته على مختلف أجزاء المحتوى وعناصره من جهة، وعلى جميع الأهداف التعليمية المتصلة بهذا المحتوى من جهة أخرى. ونقدّم فيما يلي مثالاً لجدول مواصفات اختبار تحصيلي مقنّن في اللغة العربية للصف الثالث الإعدادي أعدّه المؤلف بالاشتراك مع الدكتور محمود السيد.

(السيد ومخائيل، 1989، ص14)

الجدول (4): جدول مواصفات الاختبار (للصف الثالث الإعدادي):

موضوعات المحتوى	الأهداف التعليمية								
	المعرفة	فهم المفردات	فهم التراكيب	التعليل	استخلاص المعنى العام	استخلاص المغزى	التطبيق	التذوق (تقويم طريقة التعبير)	عدد الأسئلة
الموضوع الأول					1	1		4	6
الموضوع الثاني					1	1		2	4
الموضوع الثالث			2		1	1			4
الموضوع الرابع			4		2	2		2	10
الموضوع الخامس		6			4			2	12
الموضوع السادس		4	2		1	1			8

موضوعات المحتوى	الأهداف التعليمية								
	المعرفة	فهم المفردات	فهم التراكيب	التحليل	استخلاص المعنى العام	استخلاص المغزى	التطبيق	التذوق (تقويم طريقة التعبير)	عدد الأسئلة
الموضوع السابع		4	2		2				8
الموضوع الثامن		4	4					1	9
الموضوع التاسع			4		1	1			6
الموضوع العاشر			4		2			2	8
الموضوع الحادي عشر	11								11
الموضوع الثاني عشر							18		18
الموضوع الثالث عشر							8		8
الموضوع الرابع عشر							22		22
الموضوع الخامس عشر				16					16
الموضوع السادس عشر	12						4		16
المجموع	23	18	22	16	15	7	52	13	166

اختيار شكل البنود:

هناك عدّة أشكال أو أنواع من البنود التي يمكن استخدامها في الاختبارات الصفية. ويمكن تصنيف هذه الأنواع أو الأشكال إلى صنفين رئيسين، وهما : الأسئلة المقالية أو الإنشائية التي تتطلّب عادة إجابة طويلة مفصلة، والأسئلة أو البنود الموضوعية من نوع صواب – خطأ أو الاختيار من متعدّد أو المطابقة أو تعبئة الفراغات. وما من شك في أن اختيار النوع الملائم من البنود يعتمد إلى حد كبير على الهدف المراد قياسه، ويتعيّن على المعلّم نفسه أن يحدّد هذا النوع بما يتلاءم مع الهدف من جهة، وطبيعة المحتوى من جهة ثانية. ومن المعلوم أن بعض البنود أكثر ملاءمة من بعضها الآخر لقياس أهداف معينة. فمثلاً إذا كان الهدف هو: «أن يصبح التلميذ قادراً على تنظيم الأفكار وصياغتها بصورة منطقية» فليس من

المناسب استعمال أسئلة الاختيار من متعدد، ومن الأجدى استعمال أسئلة المقال. وإذا كان الهدف هو «تنمية قدرة التلميذ على استدعاء الأسماء والأماكن والتواريخ والحوادث» فليس من المناسب استعمال أسئلة المقال، وسيكون من المفيد استعمال أسئلة الاختيار من متعدّد أو غيرها من أنواع الأسئلة الموضوعية. ومع أن بعض الأهداف يمكن أن تقاس بأنواع مختلفة من الأسئلة فإن على المعلّم أن يختار الأفضل من تلك الأنواع. ومن الواضح أن الاختبار الواحد، قد يضم أنواعاً مختلفة من الأسئلة وذلك تبعاً للأهداف المراد قياسها كما تظهر في جدول مواصفات الاختبار. ومن الباحثين من يقترح استخدام الأسئلة الموضوعية لقياس المعرفة والفهم، واستخدام الأسئلة المقالية لقياس العمليات العليا كالتطبيق والتحليل والتركيب وغيرها (Tuckman, 1977, p.77). وليس من الصعب إثبات أن الأسئلة المقالية كثيراً ما تقتصر على قياس المعرفة فقط أو المعرفة والفهم، وتعجز عن الكشف عن مستويات التعلّم العليا، كما أن الأسئلة الموضوعية وبخاصة أسئلة الاختيار من متعدّد، إذا أحسن إعدادها، يمكن أن تتصدّى لقياس مستويات عليا في التعلّم والنشاط العقلي المعرفي. ويمكن القول: إن لكل منهما مزاياه الخاصة ضمن شروط خاصة، وإن كلاً منهما قد يصلح لتلبية أغراض معيّنة بصورة أفضل من الآخر إذا روعيت فيه تلك الشروط. وبدلاً من المفاضلة بينهما قد يكون من الأجدى العمل على استعمال كل منهما بالشكل الأفضل وتحقيق الفائدة القصوى من مزاياه الخاصة.

تقدير عدد البنود:

على واضع الاختبار أن يتّخذ قراراً حول مجموع عدد البنود التي سيضمّها الاختبار. وثمّة عوامل عديدة يجب أخذها بالحسبان عند تحديد العدد الكلي للبنود بينها:



1. نوع أو شكل البنود: فالأسئلة التي تتطلّب من التلميذ أن يضع أو يؤلف الجواب بنفسه كأسئلة المقال الطويلة أو القصيرة، وأسئلة التكميل تحتاج إلى وقت أطول من أسئلة صح – خطأ، أو الاختيار من متعدّد، أو المطابقة التي تتطلّب عادة وضع إشارة أو رقم عند الإجابة الصحيحة.
2. الغرض من الاختبار: فإذا كان الاختبار سيغطّي وحدة دراسية صغيرة فإنه يحتاج إلى عدد من البنود أقل ممّا يحتاجه اختبار فصلي أو نصف فصلي. وإذا كان الغرض من الاختبار هو تشخيص جوانب القوة والضعف في أداء المفحوص في مجال محدد لمعرفة أو مهارة فإنه يتطلّب عدداً كبيراً من البنود لتغطية المجال المراد تشخيصه بدرجة كافية، وإذا كان الاختبار مسحياً تنخفض عادة نسبة تمثيل عيّنة البنود للمجال الذي يتصدّى له الاختبار.
3. عمر التلاميذ ومستوى القدرة لديهم: فتلاميذ المرحلة الابتدائية الذين لم يتمكّنوا من مهارات القراءة والكتابة بصورة كافية يحتاجون إلى وقت أطول ممّا يحتاجه التلاميذ الكبار للإجابة عن البنود الاختبارية. والتلاميذ من ذوي القدرات المتفوقة يستطيعون الإجابة عن عدد أكبر من الأسئلة إذا قورنوا بالتلاميذ من ذوي القدرات الأدنى ممّن هم بالعمر نفسه والصف نفسه.
4. درجة الثبات المطلوبة: فكلما ازداد عدد البنود ارتفع مستوى ثبات الاختبار إذا تعادلت الشروط والعوامل الأخرى المؤثرة في الثبات. ويترتب على المعلم أن يكثر عدد البنود إذا كان يسعى إلى تحقيق درجة عالية من ثبات الاختبار على أن لا يكون ذلك على حساب شروط هامة أخرى كالصدق والقابلية للاستعمال.
5. نوع العمليات العقلية التي تتطلبها البنود الاختبارية: فالبنود التي تتطلب مجرد الاستدعاء البسيط للحقائق والمعلومات يمكن الإجابة عليها بسرعة أكبر من تلك التي تتطلب تطبيق المبادئ والمفاهيم في ظروف وأوضاع جديدة أو غير ذلك من العمليات العقلية العليا.

وعموماً يمكن القول: إنه ليس ثمة قواعد ثابتة يمكن استناداً إليها تحديد عدد البنود التي يجب أن يضمها الاختبار، ويعود هذا الأمر إلى تقدير المعلم نفسه الأدرى من غيره بخصائص تلاميذه وإمكاناتهم. وفي كل الأحوال يجب أن يكون عدد البنود كبيراً بدرجة كافية لتمثيل مجالات المحتوى والأهداف التعليمية كافة التي صمم الاختبار لقياسها.

تقدير مستوى صعوبة البنود:

إن تقدير المستوى الملائم لصعوبة البنود يمثل الخطوة الأخيرة في عملية التخطيط لبناء الاختبار. ويتحدد مستوى الصعوبة بالغرض من الاختبار. فإذا كان الغرض من الاختبار هو قياس الإتقان للأساسيات في مجال ما فيتوقع أن تكون الأسئلة سهلة نظراً لأن الطلاب جميعاً أو أكثرهم يفترض أن يكونوا قد أتقنوها. وإذا كان الاختبار تشخيصياً تميل بنوده عادة إلى السهولة نظراً لأن الاهتمام لا يتركز بتوزيع المفحوصين وفقاً لمستويات القدرة بل بالكشف عن نقاط الضعف والقوة في أدائهم. وإذا كان الاختبار من نوع الاختبارات القبلية السابقة للتدريس يتوقع أن يكون الأداء ضعيفاً لدى أغلبية الطلاب لأنهم لم يتعلموا المادة التي يغطيها الاختبار. وعموماً فإننا في اختبارات الإتقان والاختبارات التشخيصية وكذلك القبلية نكون أقل اهتماماً بالصعوبة لأننا لا نسعى إلى تحقيق انتشار واسع للدرجات، والتمييز أو التفريق بدقة بين مستويات التحصيل كما هو الحال في الاختبارات المعيارية المرجع التي تجري لأغراض الانتقاء والتصنيف والتي نسعى من خلالها إلى إنتاج أعلى تمايز بين المفحوصين وتحقيق انتشار واسع للدرجات بهدف الكشف عن الفروق الدقيقة في التحصيل. ويميل معظم علماء القياس إلى اعتبار المستوى المناسب لصعوبة البنود بنسبة 50٪ في الاختبارات التي تسعى إلى تحقيق أكبر انتشار ممكن في درجات المفحوصين، كما تسعى إلى تحقيق درجة عالية من الثبات. ومن الباحثين من يوصي بأن يبدأ الاختبار ببعض البنود السهلة بهدف استثارة الدافعية لدى المفحوصين، وحفز الضعاف منهم بصورة خاصة، وإعطائهم شيئاً من الثقة

بقدراتهم، على أن يتضمن الاختبار بنوداً أخرى صعبة تتحدى الأقوياء، مع مراعاة ان تكون أكثر الأسئلة من مستوى متوسط في الصعوبة (بنسبة 50٪). وهذا الاقتراح الأخير ينطوي على فائدة كبرى للمعلم لأن من الصعب وضع اختبار تكون بنوده جميعاً من مستوى صعوبة متوسط (بنسبة 50٪)، كما أن المعلم لا يلجأ عادة إلى التطبيق التجريبي للاختبار للتحقق من صعوبة البنود وتحديد مستواها بصورة كمية ودقيقة.

وصفوة القول: فإنه لابد من تحديد المستوى الملائم لصعوبة البنود وصعوبة الاختبار ككل بصورة مسبقة، وانطلاقاً من الغرض الخاص الذي يتوخاه المعلم من الاختبار. ويشير مهرنز في هذا الصدد إلى أن المعلم « حين يعطي بنوداً سهلة جداً أو صعبة جداً فكأنه يقول: إنه يرفض أو إنه غير قادر على إعطاء اختبارات ملائمة لطلابه في مرحلة معينة من مراحل نموهم» (Mehrens,1973,p,195) وما من شك في أن المستوى الملائم لصعوبة البنود يرتبط ارتباطاً وثيقاً بشكل هذه البنود أو نوعها. ففي اختبارات صواب – خطأ مثلاً يستحيل قبول نسبة 50٪ من الإجابات الصحيحة، وعدها دليلاً على مستوى الصعوبة الملائم إلا إذا طبقت معادلة التصحيح من أثر التخمين. ويقترح ثورندايك الاسترشاد بالجدول التالي كدليل مناسب لواضع الاختبار في محاولته كتابة اختبارات مؤلفة من أنواع مختلفة من البنود:

الجدول رقم (5): المستوى الملائم لصعوبة الأسئلة:

نوع البند	متوسط الصعوبة (نسبة الإجابات الصحيحة)
أسئلة التكميل والأسئلة ذات الإجابة القصيرة	٪50
أسئلة الاختيار من متعدد (خمسة بدائل)	٪70
أسئلة الاختيار من متعدد (أربعة بدائل)	٪74
أسئلة الاختيار من متعدد (ثلاث بدائل)	٪77
أسئلة الصواب والخطأ	٪85

إن هذه النسب تتيح الفرصة لحصول استجابات صحيحة بالتخمين، ولا علاقة لها بالنجاح والرسوب. فعملية وضع علامات أو تقديرات النجاح والرسوب مختلفة تماماً، وتميل هذه النسب برأي ثورندايك "إلى إنتاج مجموعة من العلامات تعطي أكبر فائدة للمعلم الذي يرغب في التفريق بين مستويات طلبته" (ثورندايك، 1986، ص203).

إعداد البنود الاختبارية:

تأتي بعد مرحلة التخطيط الأولية للاختبار المرحلة الثانية في عملية بناء الاختبار وهي مرحلة كتابة وإعداد البنود الاختبارية. وتنطوي هذه المرحلة على قدر كبير من الأهمية لأنها تشهد تنفيذ الخطة الموضوعة لبناء الاختبار ببنوده وتعليماته و"براعة" واضعه على مسرح الوجود. وقبل التعريف بالأنواع المختلفة من البنود الاختبارية والوقوف عند جوانب القوة والضعف فيها لابد من ملاحظة أن عملية كتابة وإعداد البنود الاختبارية هي فن لا يتم اكتسابه بصورة آلية من مجرد إتقان المادة الدرسية من جانب المعلم، وصياغة نواتج التعلم المرغوبة بصورة واضحة ودقيقة ، وفهم خصائص المفحوصين ومستويات القدرة لديهم، مع ان هذه جميعاً تمثل متطلبات أو شروطاً أساسية لهذه العملية. والواقع أن القدرة على وضع وتطوير بنود اختبارية عالية الجودة تتطلب، بالإضافة إلى ذلك كله، معرفة بمبادئ وتقنيات تصميم الاختبار ومهارة في تطبيقها، وتنطوي في كل الأحوال على عنصر الإبداع أو "لمسة خفيفة" منه. ولعل خير ما نستهل به الحديث عن عملية إعداد البنود العبارة التالية لويزمان التي قال فيها: " إن كتابة البنود هي عمل خلاق - إنها فن. وكما انه يستحيل وضع قواعد لإنتاج قصة جيدة أو رسم جيد كذلك لا يمكن وضع أحكام أو قواعد تضمن إنتاج بنود اختبارية جيدة. وبالطبع يمكن تقديم مبادئ ومقترحات، ولكن تطبيق هذه المبادئ والمقترحات يخضع لحكم واضع البنود ويحدد درجة جودتها وفاعليتها. وعند كتابة كل بند تظهر مشكلات جديدة وفرص جديدة".

(Wesman, 1971 ,p. 81)

تحليل بنود الاختبار:

بعد تصحيح الاختبارات وتثبيت الدرجات يعمد المعلمون عادة إلى إتلاف الأوراق الامتحانية أو وضعها على الرف مما يشكل خطأ كبير من وجهة نظر التربية الحديثة ونظرية التقويم المعاصرة. والواقع أن إتلاف الأوراق الامتحانية أو وضعها على الرف تمهيداً لإتلافها بعد فترة من الزمن، يؤدي إلى خسارة الجهد الكبير الذي بذله المعلم في عملية التخطيط للاختبار وبنائه، كما يحرم المعلم من فرصة تقدير فعالية البنود التي يضمها الاختبار، ويغلق أمامه بالتالي الطريق نحو بناء ملف أو مصرف شخصي للأسئلة يمكن ان يحتوي على البنود العالية الجودة التي يمكن اختيارها وتخزينها لأغراض الاستعمال في المستقبل.

إن عملية تحليل البنود تشمل مجموعة من الإجراءات منها تحديد مستوى الصعوبة (والسهولة) لكل بند، والكشف عن قدرته التمييزية بين الضعاف والأقوياء تحصيلاً، كما تشمل دراسة فعالية المشتتات أو المموهات في أسئلة الاختيار من متعدد.

ومن الفوائد والخيرات التي تقدمها عملية تحليل البنود:

1. إن تحليل بنود الاختبار ونتائجه يقدم معلومات تشخيصية تفيد في بيان مدى تعلم الصف بمجموعة أو مدى فشله في التعلم. ومثل هذه المعلومات مفيدة للمعلم وللطلاب أنفسهم، حيث أنها تلقي الضوء على مجالات الضعف العامة في الصف، كالضعف في تطبيق المهارات الحسابية أو معرفة المفردات أو فهم المبادئ... إلخ مما يتيح تركيز العمل مباشرة على تلك المجالات. ومن الواضح أن مناقشة البنود المختلفة وبخاصة الصعبة منها تفيد في تصحيح الأخطاء وإزالة سوء الفهم المؤدي لاختيار مموهات معينة.
2. يقدم تحليل البنود الأساس لتحسين التعليم الصفي من حيث أنه يكشف مدى ملاءمة المحتوى الدراسي والأهداف التعليمية للدارسين. فالمادة

السهلة جداً أو الصعبة جداً يمكن أن تؤخذ دليلاً على ضرورة إعادة النظر بالمنهاج، والأخطاء المنتشرة بين الدارسين قد تشير إلى الحاجة لإجراءات تعليمية أكثر فعالية (Gronlund, 1971,p.250).

3. ينمي المهارة في إعداد البنود. فالتحليل يكشف عن جوانب الغموض والتلميحات أو الإيحاءات، والموّهات غير الفعالة. وهذه المعلومات يمكن الإفادة منها مباشرة في مراجعة بنود الاختبار وتعديلها وتحسينها بهدف استعمالها في المستقبل. وباختصار فإن الوقوف عند الإجابات وما يتبع ذلك من مراجعة وتنقيح للبنود الاختبارية يوفر الأساس لإعداد اختبارات أفضل في المستقبل.

إجراءات تحليل البنود:

بين الإجراءات الهامة المتبّعة في تحليل البنود استخراج معاملات السهولة والتمييز لتلك البنود ودراسة فعالية "المشتتات" في أسئلة الاختيار من متعدد. ويفيد استخراج معاملات السهولة والتمييز في التأكد من ملاءمة كل من البنود التي يضمها الاختبار من حيث السهولة (أو الصعوبة)، وقدرته على التمييز بين الفيئتين العليا والدنيا من المفحوصين، كما أن دراسة "المشتتات" في أسئلة الاختيار من متعدد تتيح دراسة فعالية كل من المشتتات (الموهات)، وبيان ما إذا كانت تعمل بالاتجاه المتوقع لها. ومن الواضح أنه في ضوء هذا التحليل يمكن القيام بعملية "تنقية" شاملة للاختبار تقوم على استبقاء البنود الملائمة، واستبعاد البنود غير الملائمة، وإجراء التعديلات اللازمة على بعض البنود. ولا تتطلب عملية تحليل البنود سوى القيام ببعض العمليات الإحصائية البسيطة كما أشرنا سابقاً.

ملاحظات حول عملية تحليل البنود في الاختبار الصفي:

على الرغم من الفوائد العديدة التي تقدمها عملية تحليل بنود الاختبارات الصفية، فإن لهذه العملية حدودها ومحاذيرها التي لابد من التنبه إليها. ومن المفيد للمعلم الاطلاع على الملاحظات التالية بشأن هذه العملية:

1. القدرة التمييزية للبند لا تشير بالضرورة إلى صدقه. ومن المعلوم أن العلامة الكلية للاختبار تستخدم كأساس لاختيار الفئة العليا والفئة الدنيا ويتم تحليل كل بند في ضوء علاقته بالعلامة الكلية للاختبار. ومن هذه الزاوية فإن تحليل البنود يقدم دليلاً حول الاتساق الداخلي للاختبار أكثر من صدقه، حيث نحدد عن طريق هذا التحليل ما إذا كان كل بند يقيس بفعالية ما يقيسه الاختبار ككل، أي يعمل أو يسير بالاتجاه نفسه الذي يسير به الاختبار. وهذا يعني أن البيانات المتحصلة من تحليل البنود لا يمكن عدها بحد ذاتها دليلاً على الصدق. "وهذه البيانات يمكن أن تدعم الصدق وتؤكده حيث يتم إثبات صدق الاختبار الكلي بطريقة ما أو يفترض هذا الصدق منطقياً" (Gronlund , 1971).
2. إن انخفاض مؤشر القدرة التمييزية للبند لا يدل بالضرورة على بند سيء، فالاختبار الصفي كثيراً ما يوضع لقياس نواتج تعليمية مختلفة (المعرفة، الفهم، التطبيق... إلخ) والبند أو مجموعة البنود التي تمثل مجالاً أو هدفاً لم يأخذ وزنه في الاختبار ستكون قدرتها التمييزية ضعيفة، فمثلاً إذا كان الاختبار يضم (50) بنداً لقياس معرفة الحقائق الخاصة و (5) لقياس الفهم و(5) لقياس التطبيق، فإن البنود الأخيرة يتوقع أن يكون تمييزها ضعيفاً أي أنها لن تعمل بالاتجاه نفسه الذي يعمل به الاختبار، نظراً لأن تمثيلها في العلامة الكلية للاختبار أقل من تمثيل بنود المعرفة. وبعبارة أخرى فإن هذه البنود تقيس ما لا يقيسه الجزء الأكبر من الاختبار. ويؤدي استبعاد هذه البنود إلى جعل الاختبار مقياساً أكثر تجانساً لنواتج المعرفة، ولكنه يضر بصدق الاختبار.

3. من المعلوم أن القدرة التمييزية للبند ترتفع كلما اقترب معامل سهولته من 0,50 ، وتنخفض كلما ابتعد صعوداً أو هبوطاً عن هذا المستوى. فالبنود السهلة والسهلة جداً تضعف قدرتها التمييزية بين الضعاف والأقوياء من المفحوصين، وكذلك الأمر بالنسبة للبنود الصعبة والصعبة جداً. إلا أن إبقاء هذه البنود قد يكون أمراً ضرورياً لضمان قياس عينة ممثلة لنواتج التعلم والمحتوى الدراسي. وقد يكون لاستبعادها بهدف تعزيز الصدق مفعول عكسي حيث يؤدي إلى إضعاف مستوى الصدق بدلاً من تعزيزه لضعف التمثيل.

4. إن استقرار معاملات السهولة والتمييز للبند الاختباري، يعتمد على حجم وبنية المجموعات التي يطبق عليها هذا البند. ويرى مارتوزا أنه "من زاوية مثالية يجب أن تتم عملية تحليل البنود باستعمال عينة كبيرة (عدة مئات من المفحوصين) وغير متجانسة من حيث الخاصية التي صمم البند لقياسها (كمستوى المعرفة والقدرة... إلخ). وطالما أن هذا المطلب يصعب تحقيقه في المواقف الصفية، فلابد من الاعتماد على الحيطة والحس السليم عند تحليل البنود" (Martuza, 1977,p.183). وتبعاً لذلك فإن مؤشرات الصعوبة والتمييز لا يجوز عدها صفة ثابتة لا تتغير. والسؤال الأهم حول فاعلية البنود لا يرتبط بمستوى صعوبتها وقدرتها التمييزية بقدر ما يرتبط بدورها في التصدي لأهداف ونواتج التعلم الهامة. فقيمة البند الاختباري الصفي يجب أن ترتكز في نهاية المطاف على الاعتبارات المنطقية وليس المؤشرات الإحصائية وحدها.

5. يحدث شيء من المفارقة أحياناً حيث يسعى واضع الاختبار إلى تأمين صدق المحتوى وهو صدق منطقي قبلي يسبق تطبيق الاختبار، ثم يعمد بهدف تأكيد الصدق إلى تحليل البنود وهو إجراء تجريبي إحصائي بعدي يتم بعد تطبيق الاختبار. فمن أجل تأمين صدق المحتوى لابد من احتواء الاختبار على بنود معينة، وقد يؤدي تحليل البنود إلى ضرورة استبعادها جزئياً أو كلياً، فما هو الاختيار المناسب؟ لا شك أن الأولوية يجب أن

تعطى لصدق المحتوى، والإفراط أو المبالغة في عملية التحليل الإحصائي في الاختبارات الصفية قد لا تكون مجدية وفعالة في كل الأحيان. وهذا بالطبع لا يلغي أهمية التحليل الإحصائي في الاختبارات الصفية على الا يكون على حساب التحليل المنطقي.

6. عملية التحليل الإحصائي لا يجوز أن تقتصر على البنود الموضوعية ولكي تؤتي أكلها وتحقق الفوائد المرجوة منها لابد أن تشمل الأسئلة المقالية. إلا أن تحليل الإجابات عن الأسئلة المقالية تعتوره الكثير من الصعوبات التي يعود بعضها إلى عدم ثبات الدرجات، وإلى ان مستوى صعوبة كل جزء أو عنصر من عناصر الإجابة يؤثر في مستوى صعوبة الآخر نظراً لأن صحة (أو عدم صحة) إجابة التلميذ على جزء تؤثر في صحة (أو عدم صحة) إجابته على الذي يليه. ويرى أهمان وجلوك أن تحليل الإجابات على الأسئلة المقالية يمكن أن يتم بنجاح "إذا كانت الأسئلة من نوع الأسئلة القصيرة المحددة وتم تصحيحها بالطريقة التحليلية، حيث يمكن هنا النظر في كل جزء من الإجابة على حدة وحساب مستوى صعوبته وتمييزه، وحين يتعذر تقسيم الإجابة إلى أجزاء لأسباب منطقية، فلابد من التعامل مع السؤال ككل واتباع طريقةالاتساق الداخلي المتبعة في حالة الأسئلة الموضوعية" .(Ahmann and Glock,1975,P.161)

إعداد مصرف الأسئلة:

إحدى الفوائد الأهم لعملية تحليل البنود في الاختبارات الصفية هي إتاحة الفرصة للمعلم لإعداد ملف او مصرف شخصي للأسئلة، فاستناداً إلى نتائج التحليل يمكن اختيار أفضل الأسئلة وتخزينها بهدف استعمالها عند الحاجة. ويمكن اعتبار المصرف الشخصي للأسئلة بمثابة مكتبة مصغرة تحتوي على عدد كبير من البنود الاختبارية التي تتجه كل مجموعة منها إلى قياس هدف تعليمي محدد من خلال ارتباطه بعنصر ما من عناصر المحتوى الدراسي وكما هو مرسوم مسبقاً في جدول المواصفات. وتشبه عملية إعداد ملف للبنود

فتح حساب في المصرف في كثير من النواحي. ففي المراحل الاولى توضع الودائع ولا يتم السحب إلا بعد أن يتجمع رصيد كاف لصاحب العلاقة. وقد يبدو هذا العمل مجهداً وذا عائد محدود. ولكن خلال سنوات قليلة يمكن البدء باستعمال بعض البنود من الملف "وتغذيته" ببنود أخرى تعدّ حديثاً. وعندما "ينمو" الملف يمكن اختيار معظم البنود منه للاختبار الجديد الذي يعده المعلم. ويرى جرونلند أن للملف قيمة خاصة عندما نهتم بقياس العمليات والمستويات العقلية العليا "نظراً لأن إعداد بنود لقياس هذه العمليات أو المستويات يتطلب وقتاً وجهداً كبيرين. وفي كل مرة نعد اختباراً جديداً يستحيل وضع بنود جيدة لتلك المستويات. ومن هنا يكون أمامنا خياران : إما إهمال نواتج التعلم في المستويات العليا أو استعمال الملف" (Gronlund, 1971 ,p.257).

ويمكن إعداد الملف بسهولة من جانب المعلم إذا دونت البنود على بطاقات ودونت معها المعلومات الخاصة حول كل بند والمواصفات الفنية له، والمتحصلة من عملية تحليل البنود. وعادة يحتوي الملف على عدد من البطاقات يوازي عدد البنود بحيث تخصص بطاقة واحدة لكل بند مع ملاحظة أن طول البند أحياناً أو المعلومات المتحصلة عنه قد يتطلب عدة بطاقات. وتحتوي البطاقة عادة على البند (الأرومة أو السؤال والبدائل المعطاة للاختيار)، وتشير إلى الهدف التعليمي ومجال أو عنصر المحتوى الدراسي الذي وضع البند لقياسه، كما تحتوي على بيانات التحليل والتي تضم نتائج الفئة العليا والدنيا بالإضافة إلى معامل السهولة ومعامل التمييز، ويمكن أن تأخذ البطاقة الشكل التالي:

الشكل رقم (1) لبطاقة البيانات الخاصة بالبند الاختباري:

المادة:..

المحتوى الدراسي:....................................

الهدف التعليمي:......................................

السؤال (الأرومة):

البدائل:

أ. ....

ب. ....

ج. ....

د. ....

بيانات التحليل:

البدائل	آ	ب	ج	د
الفئة العليا	3	10	2	3
الفئة الدنيا	4	6	5	3

معامل السهولة:...................... معامل التمييز:......................

ويطبيعة الحال يمكن أن تأخذ البطاقة أشكالاً مختلفة، وقد تتضمن بيانات أخرى عن الاختبار وتاريخ إجرائه والصف وعدد الطلاب ورقم البند، بالإضافة إلى المعلومات السابقة جميعاً ، وذلك لثلاثة تطبيقات منفصلة للبند الاختباري كما هو موضح في النموذج رقم (2)، والذي يتضمن في واجهته الأمامية البند (أي السؤال والبدائل الخاصة به)، بالإضافة إلى الهدف الخاص

وعنصر المحتوى المقيس، ويتضمن في واجهته الخلفية المعلومات الأخرى وذلك وفق النموذج التالي: (نقلاً عن اهمان وجلوك Ahmann and Glock, 1975).

الشكل رقم (2) الواجهة الخلفية لبطاقة بيانات البند الاختباري:

الاختبار الصف التاريخ رقم البند عدد الطلاب						
البيانات	الفئة العليا	الفئة الدنيا	الفئة العليا	الفئة الدنيا	الفئة العليا	الفئة الدنيا
1						
2						
3						
4						
بلا إجابة						
معامل السهولة						
معامل التمييز						

ومن المهم الإشارة أخيراً إلى أن الفوائد التي يجنيها المعلم، وكذلك المهني المختص بإعداد الاختبارات ، من وراء استعمال النموذج السابق أو غيره من نماذج البيانات الخاصة للبند الاختباري كبيرة ومتنوعة، فتخصيص بطاقة (أو أكثر) لكل بند، وتضمينها المواصفات الفنية لهذا البند، والتي تم استخراجها من خلال تطبيق واحد أو أكثر، ثم تخزينها بهدف استعمالها عند الحاجة، يوفر على المعلم وقتاً وجهداً كبيرين حين يرغب في إعداد اختبار جديد يرغب في تضمينه بنوداً عالية الجودة تمّ التأكد من فاعليتها بالتحليل الإحصائي بالإضافة إلى التحليل المنطقي السابق والمرافق له. والبنود المثبتة على بطاقات يمكن فرزها ومقابلتها بالخلايا التي تمثلها في جدول المواصفات بسهولة مما يتيح تقدير صدق محتوى هذه البنود، كما يسهل عملية بناء الاختبار الجديد والذي قد

يتألف في جزئه الأعظم من تلك البنود. وما من شك في أنه بالعمل المتواصل وإضافة بنود جديدة يتمكن المعلم من تطوير اختبار تحصيلي بمواصفات جيدة. وبهذه الطريقة لا يضيع الوقت والجهد سدى، بل على العكس يتم توفير الكثير من الوقت والجهد ولكن على المدى البعيد، كما تطرأ تحسينات نوعية على عملية التقويم الصفي.

اختبارات التحصيل المقننة:

وصف عام للاختبار التحصيلي المقنن:

يمكن إجمال الصفات الأساسية للاختبار التحصيلي المقنن فيما يلي:

1. يتم إعداد هذا الاختبار من قبل فريق من المختصين والخبراء في المناهج الدراسية والقياس التربوي.
2. تتجه عادة مؤسسات أو مراكز متخصصة بنشر الاختبارات تمولها على الأغلب جهات رسمية.
3. يغطي مدى واسعاً من الأهداف التعليمية ومجالات المحتوى الدراسي.
4. يتم التخطيط له بعناية فائقة وتدرس بنوده بدقة، حيث يخضع للتطبيق التجريبي، وتحلل نتائجه لكي يشمل بصورته النهائية تلك البنود التي تتوافر فيها شروط الصلاحية كافة (الصدق والثبات والموضوعية) على وجه التحديد.
5. يزود هذا الاختبار في معظم الأحيان بقائمة معايير وذلك من خلال تطبيقه على عينة ممثلة من الطلاب من صف أو عمر معين. وهذه المعايير تقدم الأساس لتفسير الدرجات، وتمكن الذين سيستعملونه من المقابلة بين أداء تلاميذهم وأداء المجموعة المعيارية او عينة التعيير.
6. توضع لهذا الاختبار في كثير من الأحيان صور أو اشكال متعادلة يتم استخدامها لتلبية أغراض معيّنة.

ومن المفيد الاطلاع على التعريف التالي الذي وضعه نول للاختبار المقنن والذي يقول فيه: «إنه ذلك الاختبار الذي يعد بعناية من قبل خبراء في ضوء الأهداف أو الأغراض المتفق عليها. ويتم تحديد إجراءات تطبيقه وتصحيحه وتفسير درجاته بصورة دقيقة وواضحة ومفصلة، وبحيث لا يتأثر بمن يطبقه أو المكان الذي يطبق فيه. والنتائج تكون قابلة للمقارنة، ومعايير أو متوسطات المستويات العمرية أو الصفية يتم تحديدها مسبقاً (Noll,1970,p,5). ومن الواضح أن هذا التعريف يتضمن الصفات الأساسية للاختبار التحصيلي المقنن، إلا أن بعض الاختبارات التحصيلية كالاختبارات التشخيصية واختبارات الإتقان والعديد من أدوات التقويم الأخرى (كقائمة موني للمشكلات مثلاً) لا تزود بمعايير أو متوسطات المستويات العمرية أو الصفية، وتعد مع ذلك من نوع الأدوات المقننة. ومن المفيد إيضاح معنى التقنين لتكوين فكرة واضحة حول الاختبار التحصيلي المقنن.

معنى التقنين:

التقنين بالمعنى الواسع للكلمة يقتضي تحديد وضبط كل ما من شأنه أن يؤثر في أداء المفحوص، وبهذا المعنى يعتمد التقنين على رسم خطة شاملة وواضحة ومحددة لجميع خطوات بناء الاختبار وإجراءاته وطريقة تطبيقه وتصحيحه وتفسير درجاته، وتحديد السلوك المطلوب من المفحوص والشروط المحيطة به في أثناء تطبيق الاختبار، بالإضافة إلى وجود معايير لتفسير الدرجات.

ويميز علماء القياس بين معنيين اثنين للتقنين وهما:

أولاً: أن تكون تعليمات الاختبار، وصياغة بنوده، والزمن المخصص له، وشروط تطبيقه، وطريقة تصحيحه موحدة في كل المواقف، وبما يتيح الحصول على النتائج نفسها تقريباً في حال إعادة إجراء الاختبار. ومن هذه الناحية يعني التقنين التوحيد. ويفقد الاختبار أساسه العلمي والموضوعي إذا لم يكن مقنناً بهذا المعنى.

ثانياً: أن يخضع الاختبار للتقنين من خلال تطبيقه على عينة ممثلة للمجتمع الأصلي بهدف استخراج معايير معينة تحدد معنى الدرجة الخام التي يحصل عليها الفرد في ضوء تمركز وتشتت درجات أفراد عينة التقنين أو التعيير.

وغني عن البيان انه لا يمكن أن يكون الاختبار مقنناً بالمعنى الثاني دون أن يكون مقنناً بالمعنى الأول، وأن عملية استخراج معايير الاختبار، وهي المرحلة الأخيرة من مراحل تطويره، تحقق فوائد وأغراضاً عديدة منها : تعرف الوضع النسبي للفرد ضمن المجموعة، وتفسير الدرجة التي يحصل عليها، إلا أن هذه العملية – عملية استخراج المعايير، هي عملية كبيرة ومجهدة من الناحية التجريبية والإحصائية، وتستلزم تعاون عدد غير قليل من المختصين والخبراء، كما أنها مكلفة مادياً وتستغرق وقتاً غير قليل.

ومن المفيد الإشارة هنا أن الاختبار المقنن بالمعنى الأول، أي الذي أخضع لعملية التقنين دون أن تستخرج معاييره، هو أداة قياس علمية ودقيقة، ويكن أن يحقق أغراضاً وفوائد عديدة تتصل بالمتعلم والمعلم والإدارة المدرسية وتشخيص صعوبات الدراسة، كما يمكن استخدامه لأغراض البحث العلمي، وفي كل الأحوال التي لا يتركز فيها اهتمام الفاحص على تفسير الدرجة الخام التي يحصل عليها الفرد المفحوص وتعرف وضعه النسبي ضمن المجموعة.

الاختبار التحصيلي المقنن واختبار المعلم:

ثمة نقاط مشتركة بين الاختبار التحصيلي المقنن واختبار المعلم المعد بعناية من حيث أن كلاً منهما يرتكز على جدول مواصفات، ويتضمن تعليمات واضحة ومحددة للمفحوصين، وقد يأخذان الشكل نفسه من البنود الاختبارية وهو الاختيار من متعدد على الأغلب. وفي الوقت نفسه هناك فروق هامة بين هذين النوعين من الاختبارات يمكن إجمالها فيما يلي:

1. إن الاختبار التحصيلي المقنن يغطي عادة مدى واسعاً من الأهداف التعليمية وموضوعات المحتوى، وذلك لتقدير تعلم الطالب في عام أو أكثر على حين أن اختبار المعلم كثيراً ما يتصدى للأهداف الخاصة بالصف، وليس للأهداف «المشتركة» بين عدد كبير من المدارس، وقد يغطي وحدة دراسية صغيرة فقط.
2. يتم التخطيط للاختبار التحصيلي المقنن وإعداده بصورة أفضل من اختبار المعلم، حيث يتولى هذه العملية عادة فريق من المختصين والمتمرسين ببناء الاختبارات، ويخصص له الكثير من الوقت والجهد والمال. وقد يستغرق إعداد هذا الاختبار سنتين أو ثلاث سنوات.
3. يتمتع الاختبار التحصيلي المقنن بمواصفات فنية أعلى من تلك التي يمكن أن تتوافر لاختبار المعلم. حيث يخضع هذا الاختبار للتجريب بهدف تنقيته أو «غربلته»، وتجري تحليلات إحصائية موسعة لتحديد مستوى الصعوبة والتمييز والكشف عن فعالية المشتتات لكل بند.واستناداً إلى ذلك يتم تأسيس صدقه وثباته أو موثوقيته، ولا يخرج بصورته النهائية إلا بعد التأكد من أنه على درجة عالية من الصدق والثبات.
4. تزود الاختبارات التحصيلية المقننة عادة، باستثناء بعض مقاييس الإتقان والمقاييس التشخيصية، بمعايير إحصائية تفيد في مقارنة تحصيل تلاميذ صف أو عمر معين مع تحصيل عينة ممثلة من تلاميذ الصف نفسه أو العمر نفسه في البلد بكامله. ويفتقر اختبار المعلم إلى المعايير، وتتم المقارنة عن طريقه بين تلاميذ الصف الواحد في نطاق المدرسة الواحدة فقط.

خطوات بناء الاختبار التحصيلي المقنن:

تسير عملية بناء الاختبار التحصيلي المقنن ضمن مراحل وخطوات محددة من أهمها:

1. **تحديد الغرض من الاختبار:**

فالغرض من الاختبار لابد من تحديده مسبقاً والتعبير عنه وضوح. وثمة أسئلة عديدة تطرح في هذا السياق منها: هل سيكون الاختبار مسحياً أم تشخيصياً؟ هل سيتم إعداده بغرض الإرشاد الجمعي أم الفردي؟ هل ستكون هناك درجات فرعية أم درجة كلية فقط؟ ولأي مجتمع من التلاميذ يتم تصميمه؟ وهل هو لمجموعة أعمار متجانسة أم لعدة أعمار؟ وما محتواه العام؟ وهذه الأسئلة لابد من الإجابة عليها قبل البدء ببنائه.

2. **تحديد الأهداف التعليمية ومجالات المحتوى الدراسي المراد اختبارها:**

وبناء على ذلك يتم إعداد جداول مواصفات تعكس الوزن النسبي لكل هدف من الأهداف التعليمية وكل عنصر من عناصر المحتوى الدراسي وتعد هذه الجداول بمثابة خطة أو دليل عمل لبناء الاختبار كما ذكرنا سابقاً.

3. **وضع بنود اختبارية استناداً إلى جداول المواصفات:**

ولكي يشمل الاختبار بصورته النهائية أفضل البنود الممكنة يكون عدد البنود الموضوعية في البداية أكثر من العدد المطلوب لكي يتم اختيار أفضلها من خلال عملية تحليل البنود. ولابد أن تفحص البنود وتدرس بعناية في هذه المرحلة، وذلك من خلال عرضها على مختصين في المادة الدرسية وخبراء في القياس، وبطبيعة الحال لابد من إعداد تعليمات الاختبار بصورتها الأولية في هذه المرحلة.

4. تحليل البنود:

يتم تحليل البنود بعد تطبيقها على عينة من التلاميذ تشتق من المجتمع الأصلي للتلاميذ الذين صمم الاختبار من أجلهم. وإذا كان الاختبار مصمّماً للاستعمال على النطاق القومي فلابد من تطبيقه على عيّنة واسعة تمثّل المناطق الجغرافية المختلفة والمستويات الاقتصادية كافة. والغرض من هذا التطبيق التجريبي الأول هو الحصول على بيانات حول النقاط التالية:

- مستوى سهولة البند الاختباري.
- القدرة التمييزية للبند الاختباري.
- فاعلية كل مموّه لكل بند اختباري.
- تعادل البنود في الأشكال المتكافئة للاختبار في حال إعداد مثل هذه الأشكال.
- ملاءمة التعليمات وحدود الوقت.

وبناء على نتائج تحليل البنود يتم إقصاء أو تنقيح البنود الاختبارية غير الصالحة (كالبنود السهلة جداً والصعبة جداً والبنود ذات القدرة التمييزية الضعيفة) وتستبقى البنود الصالحة. وقد يتطلّب الأمر في بعض الحالات تطبيقاً تجريبياً ثانياً للاختبار بهدف المزيد من التحليل و«الغربلة»، وتأكيد الصدق والثبات. ومن المفيد في هذه المرحلة مقابلة البنود بعد أن تمّت «غربلتها» بجدول المواصفات، والتأكّد من أن الأشكال «المتعادلة» متعادلة فعلاً من حيث المحتوى والصعوبة. ولابد أن تضم الصورة النهائية للاختبار «تعليمات مقنّنة» للفاحصين والمفحوصين حول كيفية تطبيق الاختبار وحدود الوقت وإجراءات التصحيح.

5. التعيير:

بعد أن يأخذ الاختبار صورته النهائية يصبح جاهزاً للتعيير، حيث يتم تطبيقه على عيّنة واسعة وممثّلة للمجتمع الأصلي للتلاميذ على مستوى البلد بأسره أو منطقة جغرافية محدّدة. وبعد إجراء الاختبار وتصحيحه تبدأ عملية

استخراج المعايير وتقدّم على صورة جدول أو مجموعة جداول في الدليل المصاحب للاختبار. ومن المعايير: المكافئ الصفي، ويعني الوضع الصفي للأشخاص الذين يتمثّل أداؤهم المتوسط في علامة معيّنة، والمكافئ العمري ويستخرج بطريقة مشابهة، بالإضافة إلى معايير المئين والدرجة المعيارية، والدرجات المعيارية المعدّلة من مثل التساعي المعياري أو علامة ستانين، والجيمي المعياري أو العلامة الجيمية وغيرها. (انظر أيضاً الفصل الثاني الخاص بتصميم الاختبار النفسي وتقنينه من هذا الكتاب). ويتم في هذه المرحلة الأخيرة استخراج مؤشرات الصدق والثبات وتضمينها في الدليل المصاحب للاختبار.

ملاحظات هامة حول بنوك الأسئلة الامتحانية:

- مفهوم " بنك الأسئلة " هو أحد المفاهيم الحديثة العهد نسبياً التي ظهرت في عصر ثورة المعلومات والتكنولوجيا المتقدمة.
- الفكرة الموجهة لبنك الأسئلة تتمحور حول تخزين كم معين من الأسئلة تكون بمثابة رصيد أولي (أو مخزون اولي) لهذا البنك ثم رفد هذا البنك وإمداده (أو تغذيته) باستمرار بأسئلة امتحانية إضافية لتسهيل عملية "السحب" والاستخدام لهذه الأسئلة لاحقاً (ولإغناء هذا البنك لابد أن تكون كمية الأسئلة المودعة (أو الودائع)أكبر من كمية الأسئلة التي يتمّ سحبها واستعمالها. وإلا فإن هذا البنك قد يتعرض للإفلاس).
- مع أن مصطلح "بنك الأسئلة " يشير كما يبدو في الظاهر إلى عملية الحفظ والتخزين كما يشير إلى امكانية الإفادة من عملية التخزين هذه من خلال "سحب" الأسئلة واستعمالها ، فإن هذا المصطلح ينطوي على دلالات أخرى بالغة الأهمية لعل من أبرزها إخضاع الأسئلة لعملية تحليل شاملة قبل تخزينها وبحيث لا يتم تخزين أي بند أو سؤال إلا بعد دراسته وتجريبه ، والتأكد من مواصفاته الفنية.
- الأساس الذي يقوم عليه بنك الأسئلة يتمثل في التخزين كما يتمثل في تحليل الأسئلة الذي يسبق هذا التخزين. فاستناداً إلى نتائج هذا التحليل

يمكن اختيار أفضل الأسئلة وتخزينها بهدف استعمالها عند الحاجة (طبعاً في المراحل الأولى توضع الودائع ولايتم السحب إلا بعد أن يتجمع رصيد كاف). ويحسب جرونلند : للبنك أهمية خاصة عندما نهتم بقياس العمليات والمستويات العقلية العليا " نظراً لأن إعداد بنود لقياس العمليات والمستويات العليا يتطلب جهداً ووقتاً كبيرين. وفي كل مرة نعدّ اختباراً جديداً يستحيل وضع بنود جيدة لتلك المستويات. ومن هنا يكون أمامنا خياران : إما اهمال مستويات التعلم العليا ، أو استخدام البنك (او الملف)".

- عادة تدون البنود على بطاقات بحيث يخصص لكل بند بطاقة مع المعلومات الخاصة حول هذا البند ومواصفاته الفنية والمتحصلة من عملية التحليل (وقد يتطلب البند الواحد أكثر من بطاقة واحدة).
- تحتوي البطاقة عادة على البند (الأرومة او السؤال والبدائل المعطاة للاختيار) ، وتشيرإلى الهدف التعليمي ومجال أو عنصر المحتوى الدراسي الذي وضع البند لقياسه ، كما تحتوي على بيانات التحليل والتي تضم نتائج الفئة العليا والفئة الدنيا بالإضافة إلى معامل السهولة ومعامل التمييز (والمؤشرات الخاصة بفعالية المشتتات "أو البدائل الخاطئة" في أسئلة الاختيار المتعدد).
- من فوائد البنوك : إن تخصيص بطاقة أو أكثر لكل بند وتضمينها المواصفات الفنية لهذا البند والتي تمّ استخراجها من خلال تطبيق واحد أو أكثر ، ثم تخزينها بهدف استعمالها عند الحاجة ، يوفر على المعلم وقتاً وجهداً كبيرين حين يرغب في إعداد اختبار جديد يريد تضمينه بنوداً عالية الجودة تمّ التأكد من فعاليتها بالتحليل الإحصائي بالإضافة إلى التحليل المنطقي السابق والمرافق له.
- البنود المثبتة على بطاقات يمكن فرزها ومقابلتها بالخلايا التي تمثلها في جدول المواصفات بيسر وسهولة. وهذا ما يتيح التحقق من صدق محتوى هذه البنود كما يسهل عملية إعداد الاختبار الجديد والذي قد يتألف في جزئه الأعظم من تلك البنود.

- تتطلب عملية التحليل العقلي المنطقي والكمي الإحصائي للأسئلة وجود خبراء ومختصين في المجال... ولابد من التأكد من المواصفات الفنية لكل بند وتوافر مستلزمات الصدق والثبات بأشكال مختلفة.
- لعل الفائدة الأكبر من بنوك الأسئلة تتمثل في الإلحاح على مستويات التعلم العليا وعدم الاقتصار على الأسئلة التي تتناول مستويات التعلم الدنيا (أو الأسئلة الحفظية) ويمكن الإفادة من الأسئلة المخزنة واستخدامها على نطاق واسع للأغراض التقويمية المحضة (أو الامتحانية) ، إضافة لأغراض التشخيص التربوي وتحديد الصعوبات الدراسية التي يعاني منها الدارسون ، والأغراض التعليمية واستثارة دافعية الدارسين....الخ ،إضافة إلى امكان استخدامها محكات للاختبارات والأسئلة التي يعدّها المعلم بنفسه بهدف التحقق من صدقها باستخدام محك خارجي (أو الصدق بدلالة محك). وهذا الأمر بالغ الأهمية بحدّ ذاته ، لاسيما وأنه يضع المعلم وجهاً لوجه أمام إحدى المهمات الأهم المناطة إليه والتي تتلخص في إعداد أسئلة وبناء اختبارات تحصيلية في مجال اختصاصه على درجة عالية من الجودة ، وهذا ما يظهر للمعلم أن مهمته في بناء الاختبارات ليست مهمة ثانوية أو إضافية بل هي مهمة أساسية توازي تماما مهمته في التدريس.
- التكاليف والجهود اللازمة لإحداث بنوك الأسئلة ليست كبيرة إذا أخذت الفوائد التي تقدمها بالحسبان. هذا مع الإشارة إلى أن بنوك الأسئلة على أنواع فمنها البنك المصغر (أو الملف الشخصي للأسئلة الذي يتولى المعلم إعداده بنفسه في مجال دراسي أو أكثر كالرياضيات أو اللغة الانكليزية.. إلخ). ومنها البنوك التي يديرها خبراء ومختصون في المجال ويتولون بأنفسهم ، أو بالتعاون مع غيرهم ، دراسة الأسئلة التي ستتضمنها. وهذا النوع الأخير قد يكون بإشراف الدولة أو مؤسسات معينة وقد يغطي مجالاً دراسياً أو أكثر.

الفصل الخامس

الصدق

الفصل الخامس
الصدق

يعدّ مفهوما الصدق والثبات من أهم المفاهيم التي ترتكز عليها نظرية القياس النفسي إن لم تكن أهمها على الإطلاق. ومما يعزّز أهمية هذين المفهومين أنهما يمثّلان الشرطين الأساسيين من شروط صلاح الاختبار النفسي أيّاً كان نوعه. ومن المعلوم أن الاختبار النفسي يوضع أساساً لتلبية غرض أو مجموعة معينة من الأغراض، وليكون الأساس في اتخاذ عدد من القرارات المهمة. ولكي يؤدي الاختبار دوره على النحو الأمثل، ويلبي الغرض أو الأغراض المرسومة له، ويتيح، بالتالي اتخاذ قرارات صائبة لابد أن يقدّم معلومات واقعية وصحيحة حول السمة أو الخاصية التي توضع موضع القياس، أي يكون "صادقاً"، وأن يكون حساساً للفروق الدقيقة بين الأفراد في هذه السمة أو الخاصية، أي يكون "ثابتاً". وتتركز أهداف الفصل الحالي والفصل الذي يليه في إلقاء الضوء على هذين المفهومين ودراسة الطرائق التي يمكن اتباعها لتوفير قدر عال من الصدق والثبات في الاختبارات النفسية، والكشف عن الأهمية البالغة والمكانة التي يحتلها كل منهما في تلك الاختبارات.

معنى الصدق:

تجمع أدبيات القياس النفسي على أن مصطلح الصدق يشير أساساً إلى ما إذا كان الاختبار يقيس فعلاً ما أعدّ لقياسه، أو ما أردنا نحن أن نقيسه به. وقد أشار ن.ي.جرونلند في معرض حديثه عن الصدق إلى أن أقرب العبارات إلى هذا المفهوم وأكثرها دلالة عليه تلك العبارة التي يدلي بها الشاهد في المحكمة عادة التي تقول: "إنني أتعهّد أن أقول الحق، كل الحق، ولا شيء غير الحق". ومن التعريفات المهمة للصدق تعريف ليندكويست القائل: "إنه الدرجة التي يقيس بها الاختبار ما نريد قياسه"، وتعريف إدجرتون القائل: "إن الصدق يشير إلى المدى الذي تكون فيه أداة القياس مفيدة لهدف معين".

(نقلاً عن فرج، 1980، ص276 – 277)

ووفقاً لهذين التعريفين ولتعريفات أخرى عديدة لا يتسع المقام للوقوف عندها لابد من النظر إلى الصدق على أنه صفة نسبية لأداة القياس وليست مطلقة. وبدلاً من القول: إن الاختبار صادق أو غير صادق، من الأصوب أن نقول: إنه صادق بقدر ما. وهذا يعني بعبارة أخرى أن الصدق هو تعبير عن درجة، وأنه ليس ثمة اختبار صادق بصورة كلية أو مطلقة.

من جهة ثانية يرتبط الصدق مباشرة بالغرض أو الاستعمال الخاص الذي صمّم الاختبار من أجله، فالاختبار الذي أعدّ بغرض التنبؤ بأداء المفحوصين في فترة مقبلة قد لا يصلح لغرض التشخيص والكشف عن نقاط الضعف والقوة في أداء المفحوصين، أو لا يصلح لهذا الغرض بدرجة كافية. كما يرتبط الصدق مباشرة بالفئة او الجماعة التي صمّم لها، فالاختبار الذي أعدّ لفئة او فئات معينة من المعوقين قد لا يصلح للأسوياء، كما أن الاختبار الذي صمّم ليطبق على فئات معينة من الأسوياء، قد لا يصلح للمعوقين أو لجماعات أخرى من الأسوياء. ويتمثّل جوهر مفهوم الصدق في أن التعبيرات والصيغ العامة والمجردة قلّما تجدي في التعبير عن هذا المفهوم والكشف عن ماهيته، وبدلاً من استخدام تلك الصيغ والعبارات المجردة يستحسن ربط الصدق مباشرة بالغرض أو الاستعمال الخاص الذي صمّم الاختبار من أجله، ومراعاة ظروف الزمان والمكان وخصائص الجماعة التي سيطبق عليها.

وتجدر الإشارة إلى أن الأسماء التي تعطى للاختبارات النفسية لا تعبّر بالضرورة عما تقيسه هذه الاختبارات بالفعل. ومن هذا المنظور لا يصحّ أن يؤخذ الاسم الذي يحمله الاختبار على أنه تعبير أو انعكاس لما يقيسه هذا الاختبار، أو أنه مؤشر لصدقه. فالأسماء التي تحملها معظم الاختبارات النفسية تتّسم بالكثير من العمومية والغموض لدرجة يصعب معها تعرّف مجال السلوك الذي تغطيه هذه الاختبارات بدقة. وهذا يعني ببساطة أن السمة أو الخاصية التي يقيسها الاختبار فعلاً يتعذّر تحديدها من مجرد النظر في الاسم أو العنوان الذي يحمله هذا الاختبار، ولا بد من إخضاع هذا الاختبار لإجراءات معينة، قد يعتمد

الكثير منها على التجريب والإحصاء، لتحديد هذه السمة أو الخاصية المقيسة، والتثبّت من صدق الاختبار كما سنرى.

أشكال الصدق:

على الرغم من أن الباحثين استخدموا مصطلحات عديدة في التعبير عن الأنواع المختلفة من الصدق أو أشكال الصدق فإن التصنيف المعتمد لأنواع الصدق هو التصنيف الذي وضعته الجمعية الأمريكية لعلم النفس عام 1974 والذي جعل هذه الأنواع ضمن ثلاث فئات رئيسة، وهي: صدق المحتوى أو الصدق التمثيلي، والصدق المحكّي، والصدق البنيوي أو الافتراضي. وقد يكون من المفيد الإشارة، قبل الحديث عن هذه الأنواع، إلى أن بعض أنواع الصدق اكثر اهمية في أنواع معينة من الأدوات منها في أدوات أخرى. فصدق المحتوى مثلاً يؤدي دوراً بالغ الأهمية في اختبارات التحصيل على حين أن الصدق المحكّي يؤدي مثل هذا الدور في اختبارات القدرة والاستعداد. وبالطبع يمكن تصميم أداة القياس لتلبية أكثر من غرض واحد، وفي هذه الحالة لابد من الاستعانة بأكثر من نوع واحد من أنواع الصدق (Ahmann and Glock,1975). وهذا يعني انه في الكثير من المواقف الاختبارية من الممكن والمفيد الجمع بين الأنواع المختلفة من الصدق.

- **صدق المحتوى:**

يرتكز صدق المحتوى أو الصدق التمثيلي كما يسمّى عادة على الفحص المنظم لمحتوى الاختبار لتحديد ما إذا كان عينة ممثلة لمحتوى الموضوع ومجال السلوك الذي يراد قياسه. ومن المعلوم أن الاختبار، أي اختبار، يتّجه أساساً إلى قياس عينة من السلوك، وأن مجموعة البنود التي يضمّها يفترض أنها عينة من المجتمع الأصلي " الأكبر" من البنود التي كان يمكن أن يضمّها. والواقع أن المشكلة الكبرى التي تواجه واضع الاختبار عادة هي ما إذا كانت مجموعة البنود التي أعدّها هي عينة ممثلة فعلاً لكل البنود المحتملة أو " للمجتمع الأصلي"

للبنود التي تنطوي عليها السمة أو الخاصية موضع القياس. ومع أنه لا تتوفر لدى واضعي الاختبارات عادة قائمة محدودة يمكن أن يشتقوا منها عينة من البنود فإن مطلب المعاينة يمثّل المطلب الأول والأهم في عملية تطوير الاختبار وبنائه. ويستحيل دون تحقيق هذا المطلب ضمان أن يقيس الاختبار ما أعدّ لقياسه وتمثيل السمة أو ميدان السلوك المقيس بصورة أمينة وصحيحة. ومن الواضح أنه بقدر ما تكون عينة البنود ممثلة للمجتمع الأصلي الأكبر من البنود بقدر ما يرتفع صدق المحتوى، وبقدر ما يضعف تمثيل العينة لهذا المجتمع يضعف صدق المحتوى. فإذا أمكن تحديد المحتوى الدراسي أو المجتمع الأصلي للبنود بـ 300 كلمة في الهجاء مثلاً تعلّمها الأطفال خلال الفصل واحتوى الاختبار على 60 كلمة منها، أي بلغت نسبة تمثيل العينة للمجتمع الأصلي 20٪ أمكن القول: إن حجم العينة مناسب، وإن الاختبار صادق إلى درجة كبيرة في محتواه. بيد أن المحتوى الذي يشكّل المجتمع الأصلي للبنود يصعب بل يستحيل تحديده وحصره كمياً على النحو السابق في معظم الأحيان. ومع ذلك فإن هذا النوع من انوع الصدق يتطلب تحديداً ووصفاً دقيقاً للمحتوى أو المجال المقيس وتحليله إلى عناصره ومكوناته حتى يتمّ التأكد من أن البنود الموضوعة تغطّي الجوانب الأساسية لهذا المحتوى وبالنسب الصحيحة، ودون أن يطغى أي من هذه الجوانب على غيره أو يعطى له وزن أكبر من الوزن الذي يستحقه. ومن الواضح أن الاختبار بوصفه عينة مشتقة من مجتمع أصلي واسع ينطوي على قدر ما من خطأ العينة، وكلما ازداد عدد البنود التي يضمّها وكبرت العينة تضاءلت احتمالات الخطأ دون أن تنعدم، وأصبح الاختبار أكثر صدقاً وتمثيلاً للمجتمع الأصلي الافتراضي من البنود.

ليس ثمة صيغة كمية يمكن استعمالها للتحقق من صدق المحتوى أو تعبير عددي يمكن استعماله للدلالة عليه. فهذا النوع من أنواع الصدق يعتمد على التحليل المنطقي والمقارنة أو المطابقة بين محتوى الاختبار ومحتوى الموضوع المقيس، ولذلك يشار إليه في كثير من الأحيان بالصدق المنطقي. ثم إن هذا

النوع من الصدق لا يختص بنوع معين من أنواع المقاييس على الرغم من المكانة الخاصة التي يحتلها في مقاييس التحصيل. فالمقاييس التربوية والنفسية جميعها يجب أن تكون صادقة في محتواها بدرجة مناسبة، وأن يعكس أو يمثّل كل منها السمة أو الخاصية موضع القياس على النحو الملائم. فإذا كنا بصدد إعداد مقياس للاتجاهات مثلاً فلابد أن نعمل على تغطية سائر الجوانب والمظاهر التي هي مدار البحث وبما يتّسق مع الأهداف المرسومة، وكذلك الأمر فيما يتصل بمقياس الميول وغيره. بيد أن صدق المحتوى لا يكفي بحد ذاته لضمان الصدق في المقاييس النفسية، وقد يكون مضلّلاً إذا اقتصرت عليه هذه المقاييس نظراً لأنها "لا تتشابه تشابهاً" داخلياً كبيراً مع ميدان السلوك الذي تعدّ عينة منه ولا تعتمد على مقرر دراسي أو مجموعة موحدة من الخبرات السابقة يمكن منها صياغة مضمون الاختبار (أبو حطب، 1973، ص89). وعموماً فإن الاختبارات التي تستخدم لأغراض التنبؤ تتطلب الاعتماد على معايير أو محكات خارجية كما سنرى، ويصعب القول: إن هذه الاختبارات صادقة من مجرد فحص محتواها والتحقق من ملاءمتها للأهداف التي وضعت من أجلها. وقد يتعذّر تحديد الوظائف التي تقيسها تلك الاختبارات فعلاً دون الاعتماد على محكات عملية أو تجريبية تستخدم في التثبّت من صحة الفروض التي اعتمدت أساساً في بنائها. وتقول آنّا أنستازي في هذا الصدد "مع أن صدق المحتوى يجب أن يدخل في المراحل الأولية لبناء أي اختبار فإن تأسيس الصدق (أو التصديق) Validation النهائي لاختبارات الشخصية والاستعداد يتطلب التحقق التجريبي بإجراءات أخرى... ومن المستحيل تحديد الوظائف النفسية التي يقيسها الاختبار من تفحّص بنوده. فالاختبار الذي يقيس الاستدلال الحسابي لدى طلاب المرحلة الثانوية قد يقيس الفروق الفردية في سرعة العمليات العددية لدى طلاب الجامعة" (Anastasi,1982,p.136).

- **الصدق الظاهري أو السطحي:**

من المهم التمييز بين صدق المحتوى أو الصدق المنطقي للاختبار والصدق الظاهري أو السطحي. فهذا الأخير ليس صدقاً بالمعنى الدقيق للكلمة ولا يشير إلى ما يقيسه الاختبار فعلاً بل إلى ما يبدو ظاهرياً أنه يقيسه سواء في نظر المفحوصين أنفسهم أم في نظر غيرهم من الأشخاص غير المختصين. ومما لاشك فيه أن الاختبار لابد أن يبدو صادقاً كي يقبل المفحوصون عليه برغبة وتستثار دافعيتهم لأدائه بالحد الأقصى الذي تسمح به قدراتهم. وقد يؤدي غياب صفة الصدق الظاهري للاختبار حتى لو كان صدقه الحقيقي عالياً إلى موقف سلبي من جانب المفحوصين، واستيائهم مما ينعكس سلباً على أدائهم في الموقف الاختباري، كما قد يؤدي إلى التشكيك بالاختبار وقيمته ودوره في اتخاذ قرارات مهمة من جانب العديد من الأشخاص الذين قد يحكمون عليه من خلال مظهره وما يبدو لهم أنه يقيسه دونما تحليل دقيق لمحتواه ولما قصده واضعه أن يقيسه أو ما يقيسه فعلاً. ومن المعلوم أن مقياس فكسلر – بلفيو ظهر أساساً لتغطية الحاجة إلى أداة ملائمة لقياس ذكاء الراشدين نظراً لأن المقاييس الشائعة في ذلك الحين بما فيها مقياس ستانفورد – بينيه أعدت للأطفال وتلاميذ المدارس، واستخدمت مواد مشوّقة للأطفال والتلاميذ بصورة خاصة مما جعلها تفتقر إلى الصدق الظاهري عند تطبيقها على الراشدين.

تؤكد آنّا أنستازي أنه لا بد من توفير صفة الصدق الظاهري للاختبار حتى يكون أكثر فاعلية في المواقف العملية، ولضمان تعاون المفحوصين في الموقف الاختباري. ومن الواضح أن هذا التعاون والتجاوب يمثّل شرطاً ضرورياً لإظهار أقصى ما لديهم من قدرة، ويعدّ بالتالي، شرطاً ضرورياً لتوفير الصدق الحقيقي للاختبار. ويمكن أن يتحسن الصدق الظاهري بنظر أنستازي، بمجرد إعادة صياغة الأسئلة بحيث تبدو أكثر ارتباطاً بموقف القياس الفعلي. فمثلاً إذا أعدّ اختبار الاستدلال الحسابي للتطبيق على مجموعة من المشتغلين بالأعمال الميكانيكية فإن البنود يمكن صياغتها بحيث تبدو أكثر ارتباطاً بالعمل الميكانيكي ودون أن

يحدث اي تغيير في الوظيفة التي أعدت لقياسها (Anastasi,1982,p.136).
وقد يكون من المفيد في الحالات التي يتعذّر فيها توفير الصدق الظاهري للاختبار تهيئة المفحوصين مسبقاً وإعلامهم بأن الاختبار ربما سيبدو ضعيف الارتباط بما درسوه لأن غرضه هو الكشف عن قدرتهم على تطبيق ما درسوه مثلاً في مواقف وأوضاع جديدة. وهنا قد تضعف الآثار السلبية لغياب هذا النوع من الصدق، ولا يكون الاختبار بمثابة "صدمة"أو تجربة مرة للمفحوصين. ومهما يكن من أمر فإن الصدق الظاهري للاختبار لابد أن يكون في خدمة الصدق الحقيقي لهذا الاختبار وليس بديلاً عنه، ولابد من تغليب الصدق الحقيقي حين يتعذّر الجمع بينهما.

- **الصدق المحكّي:**

يرتكز الصدق المحكّي على دراسة الارتباط بين درجات المقياس ودرجات مقياس آخر يعتمد كمحك. ويمكن تعريف هذا النوع من أنواع الصدق على أنه الدرجة التي يترابط عندها الأداء على مقياس مع الأداء على مقياس آخر عدّ محكاً للمقياس الأول وأساساً في الحكم على صلاحيته. ويتّضح من هذا التعريف أن الصدق المحكّي يتطلب إجراءات تجريبية إحصائية ويقوم على حساب معامل الارتباط بين الدرجات التي يحصل عليها المفحوصون في المقياس الذي يراد التثبّت من صدقه، والدرجات التي يحصل عليها أولئك المفحوصون أنفسهم في المقياس الثاني الذي عُدّ بمثابة محك. وكثيراً ما وصف هذا النوع من الصدق بالصدق التجريبي والصدق الإحصائي نظراً لاعتماده على التجريب ولغة الإحصاء، وبهدف تمييزه عن الأنواع الأخرى من الصدق التي تعتمد على المقارنة والتحليل المنطقي أو تعطيهما الأولوية.

يميّز بعض الباحثين بين نوعين لهذا الصدق وهما: الصدق التنبؤي Predictive Validity، والصدق التلازمي أو المصاحب Concurrent Validity، ويميل الكثير من الباحثين في الوقت الحاضر إلى دمج هذين النوعين في زمرة واحدة أو نوع واحد وهو الصدق المحكّي. ويرى ن.جرونلند أن هذا الدمج مناسب

"نظراً لأن طريقة التثبّت من الصدق والتعبير عنه واحدة في الحالتين" (Gronlund,1971,p.83). ومن الواضح أن كلاً من الصدق التنبؤي والصدق التلازمي يعتمد على استخدام محك، ويتطلب حساب معامل الارتباط بين الدرجات المتحصّلة على المقياس الموضوع والدرجات المتحصّلة على مقياس المحكّ. والفرق بينهما يرجع إلى الوقت الذي تجمع فيه بيانات المحكّ، كما يرجع إلى غرض القياس. ففي حين يتركز الاهتمام عند استخدام إجراءات الصدق التنبؤي بالكشف عما إذا كانت درجات المقياس تنبئ عن أداء المفحوصين أو إنجازهم في المستقبل مما يستدعي المقارنة بين درجات المقياس الأول (المتنبئ) ودرجاتهم على مقياس آخر للأداء اللاحق (المحكّ)، يتجه الاهتمام في حالة الصدق التلازمي أو المصاحب إلى تقدير الأداء الراهن للمفحوصين، ويتطلب هذا الصدق المقارنة بين درجاتهم على المقياس ودرجاتهم على المحكّ التي يحصلون عليها في الوقت نفسه تقريباً. وهذا يعني، بعبارة أخرى، أن الصدق التنبؤي يمتد بإطاره الزمني إلى المستقبل ويسعى إلى التعبير كمياً عن درجة الترابط بين نتائج الاختبار والأداء اللاحق للمفحوصين والذي يتوقع أن يترابط مع السمة أو الخاصية التي يقيسها الاختبار، على حين أن الصدق التلازمي يرجع في إطاره الزمني إلى الحاضر، ويطلعنا على درجة التوافق أو الترابط بين نتائج الاختبار ونتائج اختبار آخر يفترض أنه يقيس السمة ذاتها.

يفيد الصدق التنبؤي في الكشف عن فاعلية الاختبار في التنبؤ بنتائج معينة في المستقبل، ويلائم بخاصة اختبارات القدرات والاستعدادات، وقد يستعمل في بعض الاختبارات التحصيلية التي يكون غرضها الانتقاء والتصنيف. ويستعمل الصدق التلازمي بخاصة في الاختبارات التي يكون غرضها تشخيص الوضع الراهن بدلاً من التنبؤ بالأداء اللاحق أو بنتائج معينة في المستقبل، ويصلح بصورة خاصة لاختبارات الشخصية، وقد يستعمل في اختبارات القدرات والاستعدادات. والمسألة المركزية في كلا النوعين من الصدق هي توفير المحكّ الملائم، فمقياس المحكّ يمثّل العنصر الحاسم في عملية التحقق من الصدق المحكّي بنوعيه. وهذا

المقياس لابد أن يتصف بصفات معينة ويكون صالحاً بحد ذاته كي يصلح أساساً في الحكم على صلاحية غيره. وقبل الوقوف عند الصفات المرغوب بها في مقياس المحكّ قد يكون من المفيد التعرّف على أنواع المحكّات المستخدمة في المقاييس النفسية والتربوية.

أنواع المحكّات:

1. محك التحصيل الدراسي:

وهذا المحكّ هو أحد أهم المحكّات المستخدمة في التحقق من صدق اختبارات القدرة العقلية العامة (اختبارات الذكاء). وكثيراً ما توصف هذه المقاييس بأنها مقاييس الاستعداد المدرسي حيث تتمّ مقارنة الدرجات المتحصّلة عليها بالدرجات والتقديرات المدرسية ودرجات الاختبارات التحصيلية المختلفة.

غير أن محك التحصيل الدراسي لا يقتصر على اختبارات الذكاء العام، وقد يستخدم في حالات معينة في التحقق من صدق اختبارات القدرات الخاصة والمتعدّدة، كما يستخدم، ولكن بدرجة أقل، في التحقق من صدق اختبارات الشخصية أو جوانب معينة منها. ومن نافلة القول: إن الكثير من اختبارات التحصيل وخصوصاً تلك التي يعدّها المعلم كثيراً ما تلجأ إلى هذا المحكّ ذاته في تأسيس صدقها وخصوصاً حين تتوفر اختبارات تحصيلية مقننة تتمّتع بمواصفات فنية جيدة ويمكن اعتمادها بالتالي محكات للصدق.

2. محك الأداء في برنامج تعليمي أو تدريبي متخصص:

ويستخدم هذا المحكّ في التحقق من صدق اختبارات القدرات والاستعدادات الخاصة، ويرتكز على أداء الأفراد خلال مرحلة التعليم أو التدريب المتخصص. فمثلاً الأداء في مدارس الفن والموسيقى يستخدم في تقدير صدق اختبارات الاستعداد الفني أو الموسيقي. والعديد من اختبارات الاستعدادات المهنية

يتمّ تقدير صدقها باستخدام محكات التحصيل في كليات الحقوق، والطب، وطب الأسنان، والهندسة، وغيرها. وكذلك اختبار انتقاء الطيارين فإن محكه هو أداء الطيارين في مواقف التدريب الفعلية (Anastasi,1982, p. 140).

3. محك الأداء في العمل نفسه:

وهو أفضل من محك سجلات التدريب السابق نظراً لأنه يرتكز على سجلات تتبّعية للإنجاز الفعلي في العمل. ومن أمثلته سجل الإنتاج الفعلي للفرد على شكل عدد القطع المصنعة، أو عدد الأخطاء المرتكبة في سجلات محاسب. غير أن هناك الكثير من الأعمال مثل طبيب، ومعلم، وسكرتيرة، التي لا يمكن فيها الحصول على سجل موضوعي للإنجاز أو الإنتاج ولذلك كثيراً ما يتمّ اللجوء إلى تقديرات الرؤساء التي قد تتأثر بعوامل ذاتية وقد تنحاز سلباً أو إيجاباً ولا تعطي صورة صادقة عن إنجاز الفرد (ثورندايك وهيجن، 1989، ص60).

4. طريقة الفرق المتقابلة:

وتنطوي على محك أكثر تعقيداً وأقل تحديداً من المحكّات السابقة. وتعتمد على المقارنة كأن نعمد إلى مقارنة درجات الطلاب في معهد الموسيقى بدرجات مجموعة غير منتقاة من طلبة الجامعة، وهذه الطريقة شائعة في اختبارات الشخصية أو اختبارات الاتجاهات والميول. ففي اختبارات الشخصية كثيراً ما تتمّ المقابلة بين العصابيين والأسوياء، وفي مقاييس الميول قد نلجأ إلى الجماعات المهنية المختلفة، وفي مقاييس الاتجاهات قد تستخدم الجماعات السياسية أو الدينية المتعارضة أو أي جماعات خاصة أخرى يمكن أن تعطي وجهات نظر متباينة أو متمايزة بصدد المسائل الخاضعة للدراسة.

5. طريقة الفروق الطرفية:

وتقترب هذه الطريقة من الطريقة السابقة من حيث أنها تقوم على مفهوم قدرة الاختبار على التمييز بين طرفي القدرة التي يقيسها، وتعتمد على المقارنة وحساب دلالة الفروق. غير أن المقارنة هنا تخص الفئتين العليا والدنيا من أفراد مجموعة واحدة ولا تخصّ مجموعتين من الأفراد. ويمكن إجراء هذه المقارنة بإحدى الطريقتين التاليتين:

1. مقارنة الفئات المتطرفة في الاختبار والمحكّ الخارجي: وبحسب هذه الطريقة تتمّ مقارنة الربع (أو الثلث) الأعلى في درجات الاختبار بالربع (أو الثلث) الأعلى في درجات المحكّ الخارجي، كما تتمّ مقارنة الربع (أو الثلث) الأدنى في درجات الاختبار بالربع (أو الثلث) الأدنى في درجات المحكّ الخارجي. وتجري هذه المقارنة عن طريق حساب الدلالة الإحصائية للفروق بين المتوسطات. فإذا لم تظهر فروق دالة إحصائياً عند مقارنة الربع الأعلى في درجات الاختبار بالربع الأعلى في درجات المحكّ، وإذا لم تظهر مثل هذه الفروق عند مقارنة الربع الأدنى في درجات الاختبار بالربع الأدنى في درجات المحكّ أمكن القول: إن الاختبار صادق بدلالة هذا المحكّ.
2. مقارنة الفئات المتطرفة في الاختبار نفسه: وتعتمد هذه الطريقة على المقارنة بين الفئات المتطرفة في الاختبار ذاته كأن يؤخذ الربع (أو الثلث) الأعلى من الدرجات المتحصّلة في هذا الاختبار (والذي يمثّل الفئة العليا) ويقارن بالربع (أو الثلث) الأدنى للدرجات فيه (والذي يمثّل الفئة الدنيا)، ثم تحسب الدلالة الإحصائية للفرق بين متوسطي الفئتين فإذا ظهرت هذه الدلالة عدّ الاختبار صادقاً (بدلالة الفرق بين الفئتين العليا والدنيا). وكثيراً ما تستخدم هذه الطريقة عند دراسة القدرة التمييزية للاختبار في المراحل الأولى لإعداده وبنائه.

وبالإضافة إلى المحكّات السابقة كثيراً ما يؤخذ الترابط بين درجات الاختبار التجريبي (الجديد) ودرجات اختبار آخر من نوعه أو قريب منه على أنه دليل للصدق. فمثلاً مقياس ستانفورد – بينيه كثيراً ما عُدّ محكاً للاختبارات الجماعية للذكاء. وترى أنستازي أن هذه الطريقة تفيد بخاصة حين يكون الاختبار التجريبي الجديد بديلاً مبسطاً أو مختصراً للاختبار الأصلي حيث لابد من اللجوء إلى هذه الطريقة عندئذ وعَدّ الاختبار الأصلي محكاً للاختبار الجديد .(Anastasi,1982,p.142)

صفات المحكّ:

ليس من السهل إيجاد محك دقيق وملائم، فمقاييس المحكّ جميعها جزئية ونسبية ولكل منها مشكلاته الخاصة وحدوده. والمحكّ النهائي، كما يقول ثورانديك: "لابد أن يتناول تقويماً لنجاح الإنسان في مهنته في الحياة اليومية. غير أن هذا المحكّ النهائي غير متيسر، ولابد للباحث أن يكتفي ببدائل هي في الغالب جزئية وغير كافية " (ثورانديك وهيجن، 1989، ص60). ومن هنا لابد من العمل على اختيار البديل الأفضل والأكثر ملاءمة من بين البدائل المتاحة.

ويرى الكثير من الباحثين أنه لابد من توفر الصفات التالية في مقياس المحكّ وهي:

1. **الصلة الوثيقة بالموضوع (Relevance):**

ويعتبر المحكّ وثيق الصلة بالموضوع بقدر ما يكون الأداء على مقياس المحكّ مناظراً أو ممثلاً للنجاح في العمل. غير أن تقدير هذه الصلة يخضع للنظرة الشخصية الذاتية وقد يختلف من مقدّر لآخر، وليس هناك أي دليل تجريبي يمكن أن يدلنا على كم هذه الصلة. ولا بد من الاعتماد على أحكام المختصين لتقدير الدرجة التي يكون عندها مقياس المحكّ، وهو مقياس جزئي

ونسبي بطبيعة الحال، وثيق الصلة بالمحكّ النهائي للنجاح في العمل والدراسة (ثوراندايك وهيجن، 1989، Mehrens,1973).

2. الثبات:

فمقياس المحكّ يجب أن ينطوي على قدر عالٍ من الثبات وبحيث نحصل عند تكراراه على نتائج متقاربة جداً. ومن الواضح أنه إذا تذبذبت العلامات أو التقديرات على مقياس المحكّ من وقت لآخر افتقد إلى الاتساق والدقة. ويتعذّر في هذه الحالة الوثوق به واستعماله كمحك.

3. الخلو من الانحياز:

وتعني هذه الصفة أن مقياس المحكّ يجب أن يتيح فرصاً متساوية للأفراد لإظهار أقصى ما لديهم من اداء والحصول على العلامات أو التقديرات التي يستحقونها. والأمثلة على عوامل التحيز كثيرة. فالرؤساء قد يتفاوتون في تقديراتهم لمرؤوسيهم تفاوتاً كبيراً في بعض الأحيان، وإذا عرف المعلم أن طالباً ما حصل على علامة منخفضة في اختبار القدرة أو الاستعداد فإن هذه المعرفة قد تؤثر في الدرجة التي سيعطيها للطالب. وقد يميل هذا المعلم إلى إعطاء علامات عالية لمن حصل على علامات عالية فقط في اختبار القدرة. وهذا ما يرفع الترابط بين درجات الاختبار والدرجات على مقياس المحكّ بصورة مصطنعة (Anastasi,1982). ومن الواضح انه كلما أمكن التغلب على عوامل الانحياز تضاءلت أخطاء التقدير وازدادت قيمة المقياس المعتمد كمحك.

لقد ظهر مصطلح خاص للتعبير عن احتمالات الخطأ (أو التحيزّات) التي يمكن أن تظهر عند استخدام مقياس المحكّ، ويعرف هذا المصطلح بـ "تلوّث المحكّ" Criterion Contamination. ويعود السبب في إطلاق هذه التسمية أو المصطلح إلى ان التقديرات التي يعطيها مقياس المحكّ قد تتأثر بقوة " أو تتلوّث" بفعل المعرفة السابقة للمقدّر بدرجة المفحوص في الاختبار. وهذا ما يظهر واضحاً

حين يكون المدرس أو الرئيس في العمل مثلاً على علم مسبق بأن علامة فلان متدنية للغاية، أو مرتفعة للغاية، فيتأثر بها بشكل أو بآخر، ويعطي تقديرات موازية لها (أي تقديرات متدنية إذا كانت العلامة متدنية، ومرتفعة إذا كانت العلامة مرتفعة). وتبعاً لذلك فإن الأشخاص الذين ستعتمد تقديراتهم أساساً في الحكم على المفحوصين (أي تستخدم كمحك) لا يجوز إطلاعهم بأي حال من الأحوال على درجات المفحوصين في الاختبار موضع الدراسة. وهذا يعني أن درجات الاختبار التي ستستخدم في "اختبار الاختبار" على حدّ تعبير انستازي وأوربينا، يجب المحافظة على سرّيتها وعدم تسريبها. هذا مع الإشارة إلى أنه ليس من السهل دائماً إقناع الأشخاص الذين تعتمد تقديراتهم كمحك (كالمدرسين، والرؤساء في العمل، والضباط وغيرهم) بأن الحذر في هذه الأمور هو شيء ضروري. وقد يفشل هؤلاء الأشخاص تماماً في إدراك حقيقة أن درجات المفحوصين يجب أن تبقى جانباً حتى "تنضج" البيانات المحكّية، ويصبح بالإمكان القيام فعلاً بحساب الصدق المحكّي للاختبار (Anasstasi&Urbina,1997,p.120)

4. جدوى المحكّ أو مدى تيسره (Availability):

فعند اختيار مقياس المحكّ لا بدّ من مراعاة أمور واعتبارات عملية من مثل الكلفة والزمن، ولا شك أن الاعتبارات العملية على أهميتها لا يجوز أن تعطى الأولوية أو تكون على حساب الصفات والشروط الأخرى السابقة.

جداول التوقع:

تستخدم جداول التوقع في التحقق من الصدق التنبؤي للمقياس. فبدلاً من محاولة الحصول على تنبؤات دقيقة حول الأداء المقبل للأفراد عن طريق معامل الارتباط يمكن الاستعانة بجداول التوقع والتنبؤ من خلال زمر أو فئات واسعة يعبر عنها برتب أو تقديرات من مثل "متوسط" أو " تحت المتوسط" أو "فوق

المتوسط". وتقدّم هذه الطريقة فائدة قصوى بخاصة للمعلم نظراً لبساطتها وسهولة استخدامها. ويمكن إيضاحها من خلال الجدول التالي:

الجدول رقم (6): جدول توقع لدرجات اختبار الاستعداد المدرسي ودرجات اختبار تحصيلي في الدراسات الاجتماعية:

درجات اختبار الاستعداد المدرسي	درجات اختبار الدراسات الاجتماعية			
	تحت المتوسط (دون 55)	متوسط (55 – 75)	فوق المتوسط (فوق 75)	المجموع
فوق المتوسط (فوق 115)	0	2	3	5
متوسط (95 – 115)	2	7	2	11
تحت المتوسط (دون 95)	3	1	0	4
المجموع	5	10	5	20

واستناداً إلى هذا الجدول يمكن القول: إنه لا أحد من التلاميذ ممن درجاتهم في اختبار الاستعداد فوق المتوسط (أكثر من 115) حصل على تقدير تحت المتوسط (أقل من 55 درجة) في اختبار التحصيل. كذلك يمكن القول: إنه لا أحد من التلاميذ ممن كانت درجته في اختبار الاستعداد تحت المتوسط (أقل من 95) حصل على علامة في اختبار التحصيل تزيد على المتوسط (أكثر من 75). وعدد التلاميذ الذين حصلوا على 95 أو أكثر في اختبار الاستعداد وحصلوا على درجات في اختبار التحصيل أقل من 55 كان 2 من أصل 16. ويصبح من السهل تفسير هذا الجدول إذا استعملت النسب المئوية. كما يظهر في الجدول التالي. (نقلاً عن: Ahmann and Glock, 1975, p.233)

الجدول رقم (7): جدول توقع يرتكز على النسب المئوية:

الاستعداد المدرسي		درجة اختبار الدراسات الاجتماعية			
الفئة	درجات الاختبار	عدد الحالات	تحت المتوسط ٪	متوسط ٪	فوق المتوسط ٪
فوق المتوسط	فوق 115	5	0	40	60
متوسط	95 – 115	11	18	64	18
تحت المتوسط	تحت 95	4	75	25	0

ومن قراءة هذا الجدول يمكن القول مثلاً: إنه إذا حصل التلميذ على تقدير متوسط في اختبار الاستعداد المدرسي فإن احتمال حصوله على تقدير تحت المتوسط في اختبار التحصيل هو 18٪ (أي 2 من أصل 11)، واحتمال حصوله على تقدير متوسط في اختبار التحصيل هو 64٪ (أي 7 من أصل 11)، واحتمال حصوله على تقدير فوق المتوسط في اختبار التحصيل هو 18٪ (أي 2 من أصل 11)، كما يمكن القول أيضاً إنه من فئة فوق المتوسط في اختبار الاستعداد المدرسي (وعدد أفرادها 5) حصل 40٪ على تقدير متوسط في اختبار التحصيل و60٪ على تقدير فوق المتوسط فيه0 ومن فئة المتوسط في اختبار الاستعداد المدرسي (وعدد افرادها 11) حصل 18٪ على تقدير تحت المتوسط في اختبار التحصيل و64٪ على تقدير متوسط و18٪ على تقدير فوق المتوسط، ومن فئة تحت المتوسط في اختبار الاستعداد المدرسي (وعدد افرادها 4) حصل 75٪ على تقدير تحت المتوسط في اختبار التحصيل و25٪ على تقدير متوسط في هذا الاختبار.

ولابد من الإشارة إلى أن جداول التوقع تفيد بخاصة في الاختبارات التي تستخدم لأغراض الانتقاء حيث يمكن استناداً إليها تحديد العلامة الدنيا أو الحد الأدنى للقبول. فإذا وجدنا مثلاً أن من نجح في تخصص دراسي أو برنامج تدريبي معين كان قد حصل على 75٪ فما فوق في امتحان القبول تثبّت هذه العلامة كحد أدنى للقبول. ويمكن القول في هذه الحالة إن امتحان القبول يتصف بالصدق التنبؤي إذ يمكنّنا من التنبؤ بنجاح من يحصل على 75٪ فما فوق ورسوب من يحصل على أدنى من ذلك.

ملاحظات مهمة حول دراسة الصدق المحكّي للاختبار:

تثير دراسة الصدق المحكّي للاختبار عدداً من الأسئلة والاستفسارات التي يتعين على العامل في القياس أخذها بالحسبان. وهذا ما يتّضح من خلال الملاحظات التالية:

1. تجانس مجموعة المفحوصين:

بما أن معامل الصدق المحكّي هو معامل ارتباط بين درجات الاختبار موضع الدراسة ودرجات المحكّ فإن كل ما يؤثّر في قيم معامل الارتباط يؤثّر في قيم معامل الصدق المحكّي. ويستتبع ذلك أنه كلما كانت مجموعة المفحوصين متجانسة انخفضت قيمة معامل الصدق المحكّي، وكلما كانت متباينة وازداد تباين الدرجات التي يعطيها الاختبار ارتفعت قيمة معامل الصدق المحكّي. وتبرز مسألة تجانس مجموعة المفحوصين في الاختبارات التنبؤية التي تعتمد نتائجها أساساً في انتقاء الأفراد. ففي هذه الاختبارات يستخرج معامل الصدق التنبؤي عادةً من النتائج المتحصّلة من تطبيق الاختبار التنبؤي على مجموعة الأفراد الذين تمّ انتقاؤهم بالفعل. وهذه المجموعة "المنتقاة" تكون بطبيعة الحال أكثر تجانساً من المجموعة الأصلية الأولى التي اختيرت منها المجموعة المنتقاة (أي

المجموع الكلي من الأفراد الذين تقدّموا للالتحاق بدراسة أو عمل معين قبل أن تتمّ عملية الانتقاء).

إن دراسة الصدق التنبؤي للاختبار الذي أعدّ بغرض الانتقاء تعتمد على الأفراد المقبولين عادة. ويستخرج معامل الصدق التنبؤي من خلال حساب الترابط بين درجات أولئك الأفراد المقبولين تحديداً في الاختبار التنبؤي ودرجاتهم في الاختبار المحكّي (أو المحكّ)، في حين أن الأفراد الذين طبق عليهم الاختبار ولم يقبلوا تستبعد درجاتهم تماماً. وهذا يعني ببساطة أن معامل الصدق التنبؤي يستخرج عادة من أداء مجموعة منتقاة. وهذا ما يؤدي بطبيعة الحال إلى ضيق مدى الدرجات التي يعطيها هذا الاختبار وضعف تباينها نظراً لأن هذه المجموعة تعدّ أكثر تجانساً من مجموعة الأفراد الذين سيطبق عليهم الاختبار في المستقبل بغرض انتقاء أفضلهم.

وفي ضوء ما سبق فإنه سيكون من المفيد عند تقويم الصدق التنبؤي للاختبار استناداً إلى معامل صدقه عدم التسرع في الحكم بعدم صلاحية هذا الاختبار، أو التشكيك بفاعليته في التنبؤ بالنجاح في الدراسة أو الأداء في العمل. فقد يعود انخفاض معامل الصدق التنبؤي إلى تجانس "أو ضعف تباين" أفراد المجموعة الذين طبق عليهم هذا الاختبار واستخرج من أدائهم بالذات هذا المعامل. وقد يعود إلى عامل آخر أو أكثر من عوامل أخرى عديدة يمكن أن تؤثر في الصدق التنبؤي للاختبار وتهبط بمعامل الصدق المستخرج.

2. ثبات درجات المحكّ:

فالمحكّات التي تستخدم في دراسة صدق الاختبار قد لا تكون على درجة عالية من الثبات أو الاتساق بحد ذاتها وخصوصاً حين تعتمد على التقديرات الذاتية ونتائج الملاحظة، كتقديرات الرؤساء في العمل، أو ملاحظات المعلم اليومية لأداء التلميذ، أو درجات الامتحانات المدرسية، أو غيرها. وقد يكون من

المناسب في الحالات التي يصعب، أو يتعذّر فيها إيجاد محكات على درجة عالية من الوثوقية والثبات استخدام أكثر من محك واحد، مع التنبّه إلى أن معاملات الصدق المستخرجة بدلالة هذه المحكّات لا تحمل المعنى ذاته الذي تحمله معاملات الصدق المستخرجة بدلالة محكات أخرى على درجة عالية من الثبات، كالاختبارات المقننة مثلاً.

3. الفترة الزمنية الفاصلة بين تطبيق الاختبار وتطبيق المحكّ:

فكلما ازدادت الفترة الزمنية الفاصلة بين تطبيق الاختبار الذي يدرس صدقه وتطبيق الاختبار الثاني الذي اعتمد محكاً للاختبار الأول ازداد احتمال ظهور عوامل دخيلة يمكن أن تؤثر في الأداء الاختباري للمفحوصين، وتؤدي إلى خفض معامل الصدق المستخرج. وبالمقابل فإنه كلما تضاءلت الفترة الزمنية الفاصلة تضاءل الأثر المحتمل للعوامل الدخيلة التي يمكن أن تهبط بمعامل الصدق المستخرج. ولعل هذا الأمر بالذات هو من بين العوامل والأسباب التي تجعل معاملات الصدق التنبؤي أدنى عموماً من معاملات الصدق التلازمي، كما تجعل معاملات الصدق التنبؤي المستخرجة بعد انقضاء فترة زمنية طويلة (3سنوات مثلاً أو أكثر) أدنى عموماً من معاملات الصدق التنبؤي المستخرجة بعد مرور فترة زمنية قصيرة نسبياً (3 أو 6 اشهر مثلاً).

4. عدد المفحوصين:

ولعدد المفحوصين أثره الواضح في الاختبارات التنبؤية خاصة. فقد تطول الفترة الزمنية الفاصلة بين تطبيق الاختبار التنبؤي والمحكّ، ويؤدي ذلك إلى انقطاع بعض المفحوصين عن الدراسة، أو تسرّبهم، أو انتقالهم، أو تعرّضهم لظروف طارئة أخرى. ويجعل من الصعب، بالتالي، وربما من المستحيل، إخضاعهم للاختبار المحكّي. والواقع أنه ليس من النادر أن يجد الفاحص نفسه أمام عدد ضئيل من الأفراد الذين طبق عليهم الاختبار التنبؤي. لهذا السبب فقد يكون من

الضروري عند دراسة الصدق التنبؤي للاختبار التنبّه مسبقاً لهذا الأمر وسحب عينة من الأفراد أكبر من العدد اللازم (علام ، 2000).

تصحيح معامل الصدق:

سبقت الإشارة إلى أن معامل صدق الاختبار يتأثر مباشرة بثبات درجات المحكّ المعتمد في استخراج هذا المعامل، وانه كلما كان المحكّ المعتمد على درجة عالية من الموثوقية والثبات أمكن الوثوق بدرجة أكبر بمعامل الصدق المستخرج بدلالة هذا المحكّ.

غير أن معامل صدق الاختبار لا يتأثر فقط بثبات درجات المحكّ المعتمد بل يتأثر أيضاً بثبات درجات الاختبار ذاته. فإذا لم يكن الاختبار موضع الدراسة على درجة عالية من الثبات والاتساق يضعف معامل صدقه الذاتي، والذي يمثّل الجذر التربيعي لمعامل الثبات، وهذا ما يؤثّر سلباً في معامل صدقه المحسوب بدلالة المحكّ المعتمد. لهذا السبب يصبح من الضروري إعادة تقدير معامل الصدق بين الاختبار والمحكّ في ضوء معامل الثبات الخاص بالاختبار، إضافة إلى معامل ثبات المحكّ.

ولإجراء تصحيح شامل لمعامل صدق الاختبار في ضوء ثبات كل من الاختبار والمحكّ تستخدم المعادلة التالية:

$$\text{ر ح أ} = \frac{\text{ر س ص}}{\sqrt{\text{ر س س} \times \text{ر ص ص}}}$$

حيث يشير الرمز ر ح أ إلى الارتباط بين المكونات الحقيقية أو التباين الحقيقي في درجات المتغيرين س، ص (الاختبار والمحكّ).

$ر_{س ص}$ إلى معامل الارتباط المحسوب بين الاختبار والمحكّ.

$ر_{س س}$، $ر_{ص ص}$ إلى معاملات ثبات كل من س وص (أي الاختبار والمحكّ).

غير أن الاختبار، أي اختبار، لابد أن ينطوي على درجة ما من الخطأ، ولا يوجد اختبار ثابت تماماً. ومن هذه الزاوية فقد نكون بحاجة لحساب صدق الاختبار في ضوء وجود هذا القدر من الخطأ والاعتراف به دون اللجوء إلى تصحيح أخطاء الثبات. وفي هذه الحالة يمكننا استخدام صيغة مختصرة للمعادلة السابقة يستبعد منها معامل الثبات الخاص بالاختبار وتبقي على معامل ثبات المحكّ، وذلك على النحو التالي:

$$ر_{ح أ} = \frac{ر_{س ص}}{\sqrt{ر_{ص ص}}}$$

حيث يشير الرمز $ر_{ح أ}$ إلى لارتباط بين التباين الحقيقي في المحكّ (ح) والاختبار (أ).

ويشير الرمز $ر_{س ص}$ إلى معامل الارتباط المحسوب بين الاختبار والمحكّ.

ويشير الرمز $ر_{ص ص}$ إلى معامل ثبات المحكّ.

ولابد أن نلاحظ عند استخدام المعادلة الخاصة بتصحيح معامل الصدق أن هذه المعادلة لا تصحح أخطاء العينة، ولا أخطاء القياس التي قد يكون لها دورها في التأثير في معامل الصدق الذي نحصل عليه بحساب الارتباط بين الاختبار والمحكّ. (Guilford,1954,p.400، نقلاً عن: فرج، 1980، ص327)

- **الصدق البنيوي (الافتراضي):**

ويشير هذا النوع من الصدق، والذي يطلق عليه أيضاً صدق التكوين الفرضي، إلى الدرجة التي يقيس معها الاختبار السمة أو الخاصية التي يفترض أن يقيسها، وينطلق من أن الدرجات على الاختبار يجب أن تتنوع أو تتباين كما تتنبأ النظرية الخاصة بالسمة المقيسة أو المفهوم أو التكوين المفترض. ومن أمثلة هذه السمات أو التكوينات الفرضية (Constructs) الذكاء، والقدرة الرياضية أو العددية، والقدرة على الاستدلال، والفهم القرائي، والتفكير النقدي، والعصابية، والقلق... إلخ. والصدق البنيوي هو مفهوم شامل يتضمّن الأنواع الأخرى من الصدق ويعدّ بالتالي شرطاً مهماً للاختبارات النفسية والتربوية بأنواعها. وبدلاً من سؤال: " هل يقيس الاختبار ما يدعي واضعه أنه يقيسه" يصبح السؤال المركزي في هذا النوع من الصدق: "ما الذي يقيسه هذا الاختبار بالضبط؟" أو "ما معنى العلامات المتحصّلة على الاختبار؟" أو "هل تشير هذه العلامات إلى سمة بارزة أو بناء معين؟". ومن المعلوم أن أدوات القياس لا تمدّنا دائماً بقياسات "نقية" للسمات التي وضعت لقياسها. ومن هذه الزاوية يؤدي هذا النوع من الصدق دوراً مهماً من حيث أنه يفيد في تحديد طبيعة وقوة العوامل التي تؤثّر فعلاً في الأداء الاختباري. فعند تفحص هذا النوع من الصدق "لا ينحصر اهتمامنا بالسمة النفسية التي صمّم الاختبار لقياسها، وقد نهتم بأي عامل آخر يمكن أن يؤثّر في الدرجات، فمثلاً قد يدعي واضع الاختبار أن اختباره يقيس الاستدلال الحسابي ولكن نحن قد نتساءل عن احتمال تأثر الدرجات بالقدرة القرائية أو السرعة أو عوامل أخرى مشابهة " (Gronlund,1971,p.91).

يتطلب الصدق البنيوي اللجوء إلى الاستنتاجات المنطقية بالإضافة إلى الوسائل التجريبية والإحصائية. والواقع أن الصدق البنيوي يتضمّن أنواع الصدق كافة كما أسلفنا. والتقنيات والإجراءات الخاصة بتأسيس صدق المحتوى والصدق المحكّي يمكن أن تندرج ضمن تقنيات وإجراءات الصدق البنيوي. فمقارنة الأداء الاختباري للفرق أو المجموعات المتقابلة كالعصابيين والأسوياء هو طريقة

من طرائق التحقق من الصدق البنيوي، وهو بطبيعة الحال طريقة للتحقق من الصدق المحكّي. ودراسة الترابط بين اختبارات الاستعداد الميكانيكي والأداء في أنواع عديدة من العمل تسهم في فهمنا للسمة التي يقيسها الاختبار مما يشير إلى دورها في الصدق البنيوي، بالإضافة إلى كونها طريقة في الصدق المحكّي. ويمكن تحديد الخطوات الواجب اتباعها في عملية التحقق من هذا النوع من الصدق على النحو التالي:

1. الدراسة المنطقية للأداة وتحديد السمات النفسية المفترضة أو البنى التي يمكن أن ينطوي عليها الأداء على تلك الأداة.
2. الدراسة المنطقية للسمات أو البنى التي تم تحديدها وتكوين فرضيات قابلة للاختبار حول مدى صلتها بالأداء على تلك الأداة انطلاقاً من نظرية تطرح تلك السمة (أو السمات).
3. إجراء بحوث للتحقق من كل فرضية بالوسائل التجريبية.

فإذا أردنا على سبيل المثال التحقق مما إذا كان اختبار ما يقيس الذكاء فيمكن أن نفترض أن الدرجات على هذا الاختبار تزيد بزيادة العمر، فإذا لم تزد يمكن أن نستنتج أحد أمرين: إما أن يكون الاختبار مقياساً غير صادق للذكاء، أو أن يكون هناك خطأ ما في النظرية التي كانت بمثابة المنطلق في تصميم هذا الاختبار.

يؤكد أريكسون أن طول عملية التحقق من الصدق البنيوي وتعقدها يجب ألا يكونا مسوّغاً لتجنبها. فليس بمستغرب، على سبيل المثال، أن نجد بين الطلاب من يظهر مهارة أو موهبة خاصة في الأوضاع الطبيعية ويعجز عن إظهار الحد الأدنى من تلك المهارة أو الموهبة في الموقف الاختباري. وفي هذه الحالة يمكن الافتراض بأن الاختبار يقيس القلق (بسبب حدود الوقت المفروضة أو إجراءات الموقف الاختباري الضاغطة). وليس بمستغرب أيضاً أن نجد بين الطلاب من يظهر أداءً عالياً في المختبر والمخزن وصالة الرسم ويظهر مع ذلك أداء ضعيفاً

على اختبارات الورقة والقلم المصمّمة لقياس التحصيل في تلك المجالات. ويمكن الافتراض في هذه الحالة أن الاختبار يقيس المهارة اللفظية أكثر مما يقيس الأداء في تلك المجالات (Erickson,1979,p.p.30 – 31). وهذا كله يظهر أهمية هذا النوع من الصدق من حيث أنه يتجه إلى الكشف عما يقيسه الاختبار فعلاً ولا ينحصر بالتأكد مما أراد واضعه أن يقيسه. والواقع أن الصدق البنيوي بوصفه مفهوماً معقداً وشاملاً لم يفهم دائماً بوضوح، وقد نظر إليه بعضهم وكأنه صدق المحتوى معبراً عنه من خلال السمة أو الخاصية النفسية المفترضة ولذا اعتمدوا على التقديرات الذاتية المحضة في التثبّت منه. وتؤكد أنستازي أن الصدق البنيوي ينطوي على قدر كبير من الأهمية، وأن هذا الصدق إذا طبق بطريقة متراخية فقد يفتح الطريق لنوع من التأكيدات الذاتية غير المحققة عن صدق الاختبار، وكل من يظن أن الصدق البنيوي يتمّ حين يفسر الاختبار كمقياس لخاصية لا يمكن تحديدها إجرائياً يرتكب خطأ فاحشاً برأي أنستازي "ولكي نكتشف ما يقيسه الاختبار فعلاً لابد أن نلجأ إلى التحقق التجريبي للعلاقات بين درجات الاختبار وبيانات أخرى خارجية".

(Anastasi,1982,p.155)

طرائق التحقق من الصدق البنيوي:

ليس هناك طريقة واحدة للتحقق من الصدق البنيوي، ولابد من جمع الأدلة من مصادر مختلفة. ويمكن استعمال الطرائق الخاصة بصدق المحتوى والصدق المحكّي كدليل جزئي لدعم الصدق البنيوي ولكن هذه الطرائق لا تكفي بحد ذاتها. فالصدق البنيوي يعتمد على الاستنتاجات المنطقية المشتقة من أنواع مختلفة من البيانات ومن الطرائق المتبعة في التحقق من الصدق البنيوي:

1. الطرائق التي تعتمد على التحليل المنطقي:

وترتكز هذه الطرائق على الفحص الدقيق لبنود الاختبار والأداء الذي تتطلبه ومقابلة نتائج هذا الفحص بالنظرية أو الفرضية المعتمدة في بناء الاختبار والتفسيرات التي قد تنبع عنها.

ويرى كرونباخ أن هذا التحليل المنطقي يعدّ من أهم مصادر التوصل إلى فروض بديلة فيما يتعلق بالأداء في الاختبارات والمقاييس، فالمحكّم الذي لديه خبرة سابقة كبيرة بالأخطاء التي شابت الاختبارات السابقة يمكنه أن يكتشف جوانب الضعف في أداة القياس الجديدة. ولكن مع ذلك فإننا لا نستطيع باستخدام هذا التحليل المنطقي للمحتوى أن ندحض صدق أداة القياس، وإنما نستطيع أن نقدم فروضاً بديلة يمكن التحقق من صحتها بالأساليب الإمبريقية أو التجريبية.

من جهة أخرى، فإن من المفيد تحليل العمليات العقلية التي يستخدمها الأفراد في أدائهم أو في التوصل إلى إجاباتهم عن بنود الاختبار، لأن هذا التحليل يضيف أدلة جديدة تزيد من فهمنا للتكوينات الفرضية التي تقيسها هذه الاختبارات (علام ، 2000، ص227). ومن هذه الزاوية فإن تحليل بنود اختبار في العلوم مثلاً قد يشير إلى أن درجاته تتأثر بالمعرفة والفهم والقدرة الكمية. وتطبيق اختبار الاستدلال الحسابي على التلاميذ وجعلهم يفكرون "بصوت مرتفع" قد يشير إلى أن البنود تتطلب عملية الاستدلال المقصودة وتعدّ بالتالي مؤشراً للصدق، أو تتطلب فقط المحاولة والخطأ وتفتقر بالتالي إلى الصدق .(Gronlund,1971 ,p.90)

2. طريقة الفرق المتقابلة:

إذا كان لدينا اختبار يفترض أنه يقيس الفروق الفردية في "الميل إلى المقامرة" فيمكن أن نؤلّف مجموعة من المفحوصين المعروف عنهم أنهم يقامرون في كل مناسبة تقريباً ومجموعة أخرى من غير المقامرين. فإذا كان الاختبار حساساً للفروق فإن الأداء المتوسط لكل من هاتين المجموعتين على الاختبار سيختلف اختلافاً ذا دلالة. وهكذا إذا اتسقت النتائج مع التوقعات يكون لدينا الدليل بأن الاختبار يقيس ما افترض أن يقيسه وهو الميل إلى المقامرة، وإذا لم تتّسق النتائج مع التوقعات فيمكن أن نستنتج أن الاختبار غير صادق أو أن مفهومنا حول سمة " الرغبة في المقامرة" هو مفهوم مغلوط. ولا يكون لدينا بذلك أي أساس للادعاء بأن الاختبار حساس للفروق الفردية في السمة مدار البحث (Marthuza,1977,p.152). وبهذه الطريقة يمكن أن نتوقع أن درجات الاختبار التحصيلي مثلاً ستميّز بين جماعة متدربة وجماعة أخرى غير متدربة، كما يمكن أن نتوقع أن الطلاب الذين يحصلون على درجات عالية على مقياس القابلية الاجتماعية لديهم من الأصدقاء أكثر من الطلاب الذين يحصلون على درجات منخفضة على هذا المقياس.

3. الترابط مع اختبارات أخرى:

حيث يتوقع أن تترابط درجات الاختبار التجريبي ترابطاً موجباً مع درجات اختبار آخر يفترض أنه يقيس السمة نفسها، وأن تترابط بدرجة أقل مع درجات اختبارات تقيس سمات أخرى. فاختبار الاستعداد المدرسي مثلاً يتوقع أن يترابط ترابطاً عالياً مع غيره من اختبارات الاستعداد المدرسي، وبدرجة مقبولة مع اختبارات التحصيل، وبدرجة أقل مع اختبار الاستعداد الموسيقي.

لقد أشار كامبل (Campbell, 1960) في معرض حديثه عن صدق التكوين الفرضي إلى أن دراسة هذا النوع من الصدق يجب ألا تقتصر على مجرد

إظهار أن الاختبار يترابط ترابطاً عالياً مع متغيرات أخرى يُتوقّع (أو يجب) أن يترابط معها. ولابد أن تعمل هذه الدراسة، بالإضافة إلى ما سبق، على إظهار أن الاختبار لا يترابط ترابطاً دالاً مع متغيرات يتوقع، أو يجب، أن يختلف عنها. وقد وصفت العملية الأولى من هاتين العمليتين بـ "الصدق التقاربي للاختبار" Convergent Validation في حين أن العملية الثانية أطلق عليها "الصدق التمييزي للاختبار" Discriminant Validation. وعلى سبيل المثال إذا أعطى اختبار الاستدلال الكمي ترابطاً عالياً مع الدرجات التحصيلية في الرياضيات لدى مجموعة من الأفراد كان هذا دليلاً على صدقه التقاربي. وأما الصدق التمييزي لهذا الاختبار نفسه فلا بد من إقامة الدليل عليه من خلال إظهار ترابط منخفض وغير دال لهذا الاختبار مع درجات اختبار آخر مختلف كاختبار الفهم القرائي مثلاً، وذلك لأن متغير القدرة القرائية لا علاقة له بما يقيسه اختبار الاستدلال الكمي.

لقد اقترح كامبل Campbell وفيسك Fiske (1959) أسلوباً لدراسة الصدق التقاربي والتمييزي للاختبار اعتمد على ما يسمى "مصفوفة السمات المتعدّدة والطرق المتعدّدة Multitrait Multimethod Matrix. وقد كان الهدف من وراء إعداد هذه المصفوفة هو الكشف عن درجة ترابط الاختبار باختبارات مشابهة للتحقق من الصدق التقاربي للاختبار، وكذلك ترابطه باختبارات مختلفة للتحقق من الصدق التمييزي أو التمايزي للاختبار.

إن مصفوفة السمات المتعدّدة والطرق المتعدّدة هي أشبه ما تكون بالتصميم التجريبي المنظم الذي يوفّر أدلة عن كل من هذين النوعين من الصدق ويتطلب إعداد هذه المصفوفة قياس اثنتين أو أكثر من السمات بطريقتين أو أكثر ولإلقاء الضوء على هذه المصفوفة وطريقة إعدادها نعرض المثال الذي يسوقه كامبل وفيسك، والذي يظهر جميع الترابطات المحتملة بين الدرجات المتحصّلة حين تقاس ثلاث سمات، كل منها على حده، بثلاث طرق. وهذه السمات الثلاث المفترضة في المثال هي:

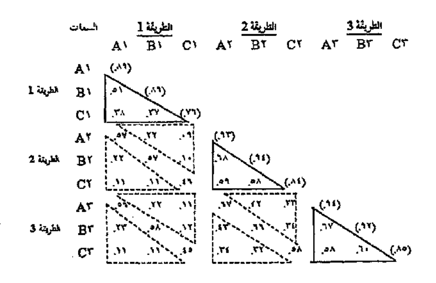
1. السيطرة.
2. الاجتماعية (أو حب الاختلاط بالآخرين) Sociability.
3. الدافعية للتحصيل.

وأما الطرق الثلاث التي يفترضها المثال فهي:

1. قائمة تعتمد على التقرير الذاتي.
2. أداة إسقاطية.
3. تقديرات الأقران.

وطبقاً لهذا المثال، والذي يعرضه الجدول رقم (7)، تشير الأحرف A ،B ،C إلى السمات الثلاث المفترضة في حين تشير الأرقام السفلية إلى الطرق الثلاث. وعلى هذا يشير A1 إلى درجات السيطرة في قائمة التقرير الذاتي، وA2 إلى درجات السيطرة في الأداة الإسقاطية، وA3 إلى درجات السيطرة كما تحددها تقديرات الأقران. كما يشير B1 إلى درجات الاجتماعية في قائمة التقرير الذاتي، وB2 إلى درجات الاجتماعية في الأداة الإسقاطية وهكذا....

الجدول رقم (8): مصفوفة السمات المتعددة والطرق المتعددة لمجموعة من البيانات الافتراضية (نقلاً عن: Anastasi &Urbina ,1997,p.131)



	السمات	A١	B١	C١	A٢	B٢	C٢	A٣	B٣	C٣
		الطريقة 1			الطريقة 2			الطريقة 3		
الطريقة 1	A١	(.٨٩)								
	B١	.٥١	(.٨٦)							
	C١	.٣٨	.٣٧	(.٧٦)						
الطريقة 2	A٢	.٥٧	.٢٢	.٠٩	(.٩٣)					
	B٢	.٢٢	.٥٧	.١٠	.٦٨	(.٩٤)				
	C٢	.١١	.١١	.٤٦	.٥٩	.٥٨	(.٨٤)			
الطريقة 3	A٣	.٥٦	.٢٢	.١١	.٦٧	.٤٢	.٣٣	(.٩٤)		
	B٣	.٢٣	.٥٨	.١٢	.٤٣	.٦٦	.٣٤	.٦٧	(.٩٢)	
	C٣	.١١	.١١	.٤٥	.٣٤	.٣٢	.٥٨	.٥٨	.٦٠	(.٨٥)

تتضمّن المجموعة الأولى من الترابطات المفترضة التي يعرضها الجدول السابق وهي الأرقام الواقعة بمحاذاة القطر الرئيسي للمصفوفة (بين الأقواس) – تتضمّن قيم معاملات ارتباط كل سمة بنفسها بعد أن أخضعت للقياس بإحدى الطرق (أو الأدوات) الثلاث (أي ارتباط A1 مع A1 وB1 مع B1 وC1 مع C1 وهكذا). وهذا يعني بعبارة أخرى أن هذه المجموعة من الترابطات والمؤلفة من تسعة ترابطات تشير إلى ارتباط كل اختبار بنفسه فيما يخصّ كل سمة، أي أنها تشير إلى معامل ثبات كل من الاختبارات التسعة (ومن الطبيعي أن تكون قيم هذه الترابطات مرتفعة لكونها تعبر عن ارتباط كل اختبار بنفسه، وتعكس مدى اتساقه مع نفسه).

أما المجموعة الثانية من الترابطات التي يعرضها الجدول السابق فتمثّل قيم معاملات الصدق وهي الأرقام التسعة الواقعة بمحاذاة الخطوط القطرية القصيرة والملونة باللون الأسود (أي الأرقام: 0.57، 0.57 ، 0.46، 0.67، 0.66، 0.58، 0.56، 0.58، 0.45). وتشير معاملات الصدق هذه إلى ترابطات الدرجات المتحصّلة في السمة ذاتها بطرائق (أو أدوات) مختلفة، وهي في مثالنا هذا تشير إلى ترابط قائمة التقرير الذاتي مع الأداة الإسقاطية، وترابط قائمة التقرير الذاتي مع تقديرات الأقران، وترابط الأداة الإسقاطية مع تقديرات الأقران في كل سمة من السمات الثلاث (أي 3 ترابطات لكل سمة و9 ترابطات للسمات الثلاث). ومن الواضح أن معاملات الصدق المستخرجة هنا توفّر الدليل عن الصدق التقاربي للاختبار، وقد كانت قيم هذه المعاملات مرتفعة نسبياً ولكن دون أن تصل إلى قيم المجموعة السابقة من الترابطات التي تشير إلى معاملات الثبات وتعبر عن ارتباط كل اختبار بنفسه كما سبقت الإشارة.

وبالإضافة إلى الترابطات السابقة يتضمّن الجدول السابق الترابطات بين السمات المختلفة وقد قيست بالطريقة نفسها (وهي القيم المدوّنة داخل المثلثات التي تحدّها خطوط متصلة)، كما يتضمّن الترابطات بين السمات المختلفة وقد قيست بطرق مختلفة (وهي القيم المدونة داخل المثلثات التي تحدها خطوط

متقطعة). ولتوفير صفة الصدق البنيوي (أو صدق التكوين الفرضي) لابد أن تكون معاملات الصدق المحسوبة أعلى بوضوح من الترابطات بين السمات المختلفة المقيسة بطرق مختلفة، كما يتوقع أن تكون أعلى (ولكن بدرجة أقل) من الترابطات بين السمات المختلفة المقيسة بالطريقة نفسها. فمثلاً الترابط بين درجات السيطرة المتحصّلة من قائمة التقرير الذاتي ودرجات السيطرة المتحصّلة من الأداة الإسقاطية (وهو ما يشير إلى الصدق التقاربي) يجب أن يكون أعلى من ترابط درجات السيطرة المتحصّلة من قائمة التقرير الذاتي ودرجات الاجتماعية المتحصّلة من الأداة الإسقاطية. في الوقت نفسه يتوقع أن يكون هذا الترابط (أي الترابط بين درجات السيطرة المتحصّلة من قائمة التقرير الذاتي ودرجات السيطرة المتحصّلة من الأداة الإسقاطية) أعلى (ولو قليلاً) من ترابط درجات السيطرة المتحصّلة من قائمة التقرير الذاتي ودرجات الاجتماعية المتحصّلة أيضاً من قائمة التقرير الذاتي.

ويتبين مما سبق أن دراسة الصدق التقاربي والتمييزي وفقاً للنهج الذي اختطه كامبل وفيسك تعتمد على دراسة الارتباط بين الطرق (أو الأدوات) المختلفة التي تقيس السمة نفسها (وهذا ما يعبر عنه بالصدق التقاربي) ثم مقارنة هذا الارتباط بالارتباط بين السمات المختلفة التي تقاس بالطريقة نفسها، وكذلك الارتباط بين السمات المختلفة التي تقاس بطرق مختلفة (وهذا ما يعبر عنه بالصدق التمييزي). وهذا يعني أن دراسة الصدق التقاربي لا تنفصل عن دراسة الصدق التمييزي، بل تصبّان في مجرى واحد، وتمثّلان جانبين لعملية واحدة متكاملة هدفها الأول والأخير هو التثبّت من صدق الاختبار باعتماد الأسلوب "التقاربي" من جهة والأسلوب التمييزي أو التباعدي من جهة أخرى.

4. التحليل العاملي:

ويهدف إلى تحديد العوامل أو السمات المشتركة بين مجموعة من الاختبارات وذلك بحساب معاملات الارتباط بين كل اختبار والاختبارات الأخرى فإذا ظهرت تجمعات معينة بينها فهذا يدل على وجود سمات أو عوامل مشتركة بينها، وهذا يعني أن السمة أو العامل المشترك نستدل عليه من وجود ترابط عالٍ بين عدد من الاختبارات. فمثلاً إذا كانت الاختبارات التي تترابط مع بعضها ترابطاً عالياً هي من نوع اختبارات المفردات وتكملة الجمل والمتشابهات والأضداد، وكانت معاملات ارتباط هذه الاختبارات ببقية الاختبارات منخفضة فيمكن أن نستنتج وجود عامل الفهم اللغوي. ومن الواضح أن التحليل العاملي يؤدي دوراً بالغ الأهمية في تقدير الصدق البنيوي من خلال الكشف عن السمة التي يقيسها الاختبار بأسلوب كمي إحصائي.

يتيح التحليل العاملي إنقاص عدد المتغيرات التي يمكن أن يوصف من خلالها أداء الفرد. فبدلاً من أن يوصف أداء الفرد من خلال 20 اختباراً أخضع لها هذا الفرد مثلاً، يمكن أن يوصف هذا الأداء من خلال خمسة أو ستة عوامل أمكن استخراجها من خلال التحليل العاملي لهذه الاختبارات وتبيّن أنها المسؤولة عن الترابطات الداخلية بينها. وعلى هذا يمكن التعبير عن أداء الفرد من خلال درجاته في خمسة أو ستة عوامل بدلاً من درجاته في الاختبارات العشرين. وكما سبقت الإشارة فإن الهدف الرئيس للتحليل العاملي هو تبسيط عملية وصف السلوك عن طريق خفض عدد المتغيرات التي تقيسها الاختبارات إلى عدد ضئيل من العوامل (أو السمات) المشتركة.

هذا ويمكن استخدام العوامل التي يتمّ استخراجها بالتحليل العاملي في وصف البنية العاملية للاختبار. وعلى هذا الأساس يمكن وصف كل اختبار بالاعتماد على العوامل الرئيسة المحدّدة لدرجاته، إضافة إلى تشبّعه بكل من هذه العوامل وارتباطه بها وهذا ما يطلق عليه الصدق العاملي Factorial validity

للاختبار. وهكذا فإذا ترابط عامل الفهم اللفظي بمقدار 0.66 مع اختبار المفردات فإن الصدق العاملي لهذا الاختبار بوصفه مقياساً لسمة الفهم اللفظي هو 0.66.

وتجدر الإشارة هنا إلى أن الصدق العاملي يشير إلى ترابط الاختبار مع كل ما هو مشترك بين مجموعة من الاختبارات، أو غيرها من الأدوات "اللااختبارية". وهذا يعني أن التقديرات الرتبية والمقاييس المحكّية الأخرى يمكن استخدامها، إضافة لاختبارات أخرى، في استطلاع البنية العاملية لاختبار معين وتحديد السمات المشتركة التي يقيسها.

(Anastasi&Urbina ,1997,p.128)

تكامل الطرائق المتبعة في دراسة الصدق:

على الرغم من أن الأشكال أو الطرائق المتبعة في دراسة الصدق تصنّف عادة إلى ثلاثة أشكال رئيسة "متمايزة"، وهي: صدق المحتوى، والصدق المحكّي، وصدق التكوين الفرضي، وأن لكل منها أساليبه وإجراءاته الخاصة، فإن هذه الطرائق أو الأشكال لا تمثّل في حقيقة الأمر أنواعاً متباينة من الصدق. وقد يكون من الخطأ النظر إليها بصورة منفصلة، أو الاعتقاد بأن أياً منها يغني عن الآخر أو يحل محله، والأصح أن ننظر إليها على أنها جوانب أو مظاهر متعدّدة لعملية واحدة متكاملة وهي عملية التحقق من الصدق بالمعنى الشامل. هذا مع الأخذ بالحسبان أنه ليس من الضروري دائما استخدام هذه الطرائق جميعها عند دراسة الصدق لاختبار محدد، وأن لكل من هذه الطرائق والأشكال استعمالاته الخاصة به، وذلك انطلاقاً من أن صدق أدوات القياس ليس مفهوماً عاماً، وإنما هو مفهوم نسبي يرتبط بمواقف، أو استخدامات معينة لمجموعات معينة من الأفراد.

إن اختيار شكل أو أكثر من أشكال الصدق عند دراسة الصدق لاختبار محدد يتوقف على الغرض أو الاستعمال الخاص الذي صمّم من أجله هذا

الاختبار. ومن هذا المنظور لابد من اختيار شكل أو أكثر من أشكال الصدق يكون أو تكون أكثر ملاءمة للغرض الذي صمّم الاختبار من أجله. في الوقت نفسه، لا يصح بأي حال من الأحوال، الاقتصار على شكل واحد من أشكال الصدق حين يكون هناك أكثر من غرض واحد للاختبار. فالاختبار الذي صمّم لتلبية أغراض عديدة ومختلفة لابد من دراسة صدقه بطرائق عديدة ومختلفة. وحين يستخدم الاختبار التحصيلي في التنبؤ بالأداء اللاحق للفرد في المرحلة الدراسية الأعلى مثلاً يصبح من المحتم دراسة صدقه باعتماد محك معين للأداء اللاحق، وعدم الاقتصار على دراسة صدق محتوى هذا الاختبار. وتظهر في الجدول التالي الطرائق (أو الأشكال) التي يمكن استخدامها عند دراسة الصدق لاختبار في الحساب أعطي كمثال، وارتباط كل منها بالغرض أو الاستعمال الخاص به (الجدول رقم (9)):

الجدول رقم (9) دراسة الصدق لاختبار في الحساب (صمّم لتلبية أغراض متعدّدة)

الغرض من القياس	السؤال التوضيحي	دليل الصدق
قياس التحصيل في الحساب "الابتدائي"	كم تعلّم التلميذ "س" من مادة الحساب؟	وصف المحتوى
اختبار للاستعداد (بهدف التنبؤ بالأداء اللاحق في الرياضيات)	إلى أي درجة يمكن أن ينجح التلميذ في المستقبل في دراسة الرياضيات؟	محك للتنبؤ
أداة لتشخيص صعوبات التعلّم	هل يشير أداء التلميذ إلى صعوبات محددة؟	محك للتنبؤ (يستخدم في الظرف الراهن)
قياس الاستدلال الكمي	كيف ترتبط درجة التلميذ بمؤشرات أخرى للقدرة الاستدلالية؟	تحديد السمة المفترضة

(نقلاً عن: Anastasi&Urbina ,1997,p.136)

لقد سبقت الإشارة إلى أن الصدق البنيوي أو الافتراضي هو مفهوم عام شامل يتضمّن أنواع الصدق كافة، وأن الأساليب والإجراءات والتقنيات الخاصة بتأسيس صدق المحتوى وكذلك الصدق المحكّي يمكن أن تندرج ضمن أساليب الصدق البنيوي وتقنياته وإجراءاته. والواقع أن البيانات والأدلة التي تجمع للتحقق من الصدق البنيوي للاختبار تتضمّن عادة من جملة ما تتضمّن أدلة عن صدق محتواه وصدقه المحكّي، مما يشير إلى أن الصدق البنيوي يعدّ الإطار الشامل الذي تلتقي ضمنه كما تتكامل من خلاله أشكال الصدق كافة. ولعل في هذه النظرة إلى الصدق البنيوي إشارة واضحة إلى تكامل الطرائق المتبعة في دراسة الصدق، وتأكيداً لدورها مجتمعة في هذه الدراسة، دون أن ينفي هذا حقيقة ارتباطها بأغراض محددة، وإمكانية الاقتصار على إحداها أو بعضها في الحالات التي يمكن أن تلبي الغرض المرسوم للاختبار.

العوامل المؤثرة في الصدق:

ثمة عوامل عديدة يمكن أن تؤثّر في صدق الاختبار وتضعفه. وقد صنف ن. جرونلند هذه العوامل ضمن أربع فئات رئيسة وهي التالية:

1. **عوامل في الاختبار نفسه:**

ومن أهمها:

أ. التعليمات غير الواضحة: فالتعليمات التي لا تشير بوضوح إلى ما يجب أن يفعله التلميذ، وكيف يجيب، وما إذا كان يسمح له بالتخمين، وكيف يسجل إجاباته، تضعف الصدق.

ب. المفردات والتراكيب الصعبة: فهي قد تجعل الاختبار يقيس الفهم القرائي أو مظاهر من الذكاء وليس مظاهر السلوك التي أعدّ لقياسها.

ج. البنود الصعبة جداً أو السهلة جداً: إذ أن البنود الصعبة جداً وكذلك السهلة جداً لا تميّز بين الأقوياء والضعاف تحصيلاً كما سنرى، وبالتالي فهي تضعف الصدق.

د. البنود الموحية بالإجابة: فمثل هذه البنود قد تقيس القدرة على اكتشاف الإجابات الصحيحة اعتماداً على الإيحاءات أو التلميحات التي تنطوي عليها البنود ذاتها ولا تكشف بالتالي الفروق في السمة أو الخاصية موضع القياس.

ه. عدد البنود: فالاختبار هو عينة من بنود أو أسئلة كثيرة يمكن أن تطرح وإذا كان قصيراً وتضمّن عدداً ضئيلاً من البنود يضعف تمثيله وبالتالي يضعف صدقه.

و. الغموض: فالغموض يؤدي إلى سوء التفسير وقد يضلّل المفحوصين الأقوياء أكثر من الضعاف ويجعل الاختبار يميّز بالاتجاه المعاكس.

ز. ترتيب البنود بصورة غير ملائمة: يرتّب الاختبار عادة وفق مبدأ التدرج في الصعوبة فإذا وضعت البنود الصعبة في البداية يقضي الطلاب الكثير من الوقت للإجابة عنها، وقد لا يتسع لهم الوقت للإجابة عن البنود السهلة. ثم إن هذا الترتيب قد يضعف دافعيتهم مما ينعكس سلباً على أدائهم.

ح. تخصيص مكان واحد للإجابة الصحيحة: فإذا أخذ الجواب الصحيح الرقم (3) أو الرمز (ج) مثلاً يسهل اكتشافه مما يؤدي إلى إضعاف الصدق بطبيعة الحال.

2. العوامل المتصلة بشروط الإجراء والتصحيح ومنها:

أ. الشروط البيئية المحيطة: فالحرارة والبرودة الزائدة والضوضاء وغيرها، هي من العوامل التي يمكن ان تؤثّر سلباً في الأداء الاختباري.

ب. الوقت المخصص للإجابة: فإذا لم يكن كافياً أو كان طويلاً أكثر من اللزوم يتأثر الأداء الاختباري بالضغط على المفحوصين في الحالة الأولى وإرباكهم، وإتاحة الفرصة للغش والفوضى في الحالة الثانية.

ج. أخطاء التصحيح: وقد تنتج هذه الأخطاء عن عملية التقدير ذاتها كما في الأسئلة المقالية التي تعاني من عيوب التقدير الذاتي وضعف مستوى الثبات، وقد تنتج عن عملية جمع العلامات أو نقلها وما شابه ذلك.

د. الغش: فإتاحة الفرصة للغش أو تقديم المساعدة للمفحوصين يحولان دون الكشف عن الفروق الحقيقية بينهم، ويضعفان الصدق.

3. العوامل المتصلة باستجابات المفحوصين:

قد يتعرض المفحوصون لأزمات انفعالية حادة تؤثر في أدائهم الاختباري. ومن المعلوم أن بعض المفحوصين يعانون بشدة مما يسمى بـ " قلق الاختبار"، كما أن بعضهم قد تضعف دافعيتهم ولا يبذلون جهودهم على النحو المطلوب، وهذا ما يظهر بخاصة في الاختبارات التي تجرى لأغراض البحث التي قد لا يعيرها المفحوصون أو بعضهم الاهتمام الكافي نظراً لأنها قد لا تعني شيئاً بالنسبة لهم أو لا تقدّم أية فائدة من وجهة نظرهم.

بيد أن العامل الأقل وضوحاً بين العوامل المتصلة باستجابات المفحوصين والذي ينطوي مع ذلك على قدر كبير من الأهمية هو ما يسمى نمط الاستجابة (Response Set) والذي يظهر في ميل ثابت نسبياً لاتباع نموذج معين في الإجابة على البنود. فمثلاً بعضهم قد يعطي إشارة "صح" باستمرار على أسئلة صح – خطأ، وبعضهم الآخر قد يعطي إشارة خطأ باستمرار بغض النظر عن صحة الجواب أو خطئه. فإذا احتوى الاختبار على عدد كبير من البنود الصحيحة يكون في مصلحة الفئة الأولى وضد الثانية والعكس بالعكس. ومع أن بعض أنماط الاستجابات يمكن مواجهتها والتغلب عليها كأن نجعل عدد الإجابات الصحيحة مساوياً عدد الإجابات الخاطئة فإن أنماطاً أخرى يصعب ضبطها والسيطرة عليها كالميل للعمل بسرعة على حساب الدقة، والميل إلى المجازفة عند الشك، واتباع أسلوب معين في الإجابة عن أسئلة المقال. فهذه الأنماط تضعف الصدق لأنها تدخل عوامل لا علاقة لها أساساً بغرض القياس (Cronbach,1970).

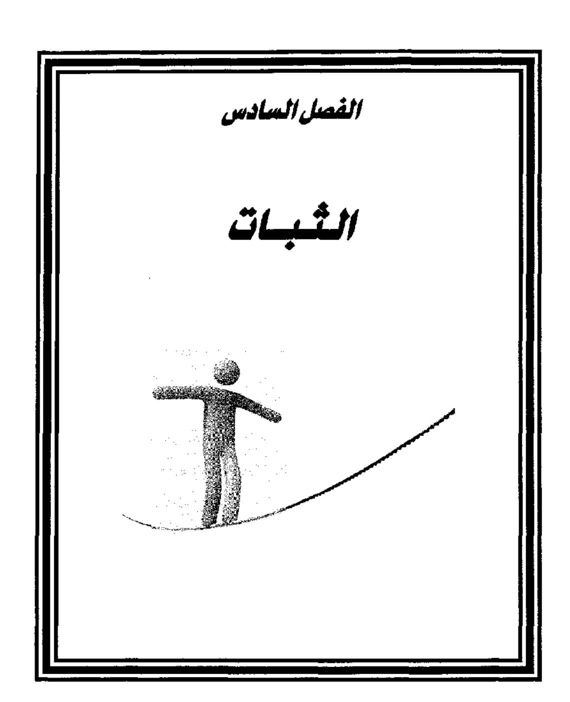
4. طبيعة المجموعة والمحكّ:

فالصدق خاص بمجموعة معينة من المفحوصين. ودرجات اختبار العلوم، على سبيل المثال، قد تتأثر بالفهم القرائي لدى مجموعة من المفحوصين ولا تتأثر بذلك لدى مجموعة أخرى لا تعاني من صعوبات في الفهم القرائي. والواقع أن ما يقيسه الاختبار يتأثر بعوامل عديدة كالعمر، والجنس، ومستوى القدرة، والخلفية التربوية، والخلقية الثقافية، وغيرها. ومن الضروري في تقويم معاملات الصدق مراعاة طبيعة المحكّ المستعمل. فنحن يمكن أن نتوقع أن تترابط درجات اختبار التفكير النقدي ترابطاً عالياً مع درجات اختبار الدراسات الاجتماعية التي تلح على التفكير النقدي أكثر من تلك التي تعتمد على تذكر الحقائق. وكلما ازداد التشابه بين السلوكات التي يقيسها الاختبار والسلوكات التي يمثّلها المحكّ ارتفع معامل الصدق.

وعموماً فإن أي عامل يؤثّر في أداء المفحوص ويمنعه عن إظهاره على حقيقته يضعف صدق الاختبار ويحرف الاختبار عن غرضه الأصلي في الكشف عن مستوى الأداء الحقيقي للمفحوص (Gronlund,1971,p.p93 – 96).

الفصل السادس

الثبـات



الفصل السادس
الثبات

معنى الثبات:

يمثّل الصدق المطلب الأساسي لصلاحية المقياس. أما الثبات فهو شرط ضروري أو لازم للصدق يستحيل دونه تأسيس الصدق وتأكيده. فلكي نقول: إن المقياس يقيس ما وضع لقياسه أو إنه ينطوي على درجة عالية من الصدق يتعين علينا التأكد مما إذا كان يقيس بدقة (ذلك الشيء) الذي وضع لقياسه. والسؤال المركزي المطروح بصدد الثبات هو: ما مستوى الدقة الذي توفّره أداة القياس؟ أو كم يتحقق فيها من الدقة والاتساق عند تكرار القياس على مجموعة من الأفراد؟ وبعبارة أخرى، هل يكشف المقياس بصورة ثابتة ومطّردة الفروق بين الأفراد، وكم تتذبذب نتائجه؟ ومن الواضح أنه كلما تذبذبت النتائج انخفض ثبات الاختبار وضعفت حساسيته للكشف بدقة عن الفروق الحقيقية في السمة أو الخاصية موضع القياس.

يستخدم مفهوم الثبات بالمعنى العام للدلالة على مدى اعتماد الفروق في درجات الاختبار على الفروق الحقيقية في السمة أو الخاصية المقيسة من جهة وأخطاء القياس العشوائية أو الناجمة عن الصدفة من جهة أخرى. ومن المعلوم أن الاختبار في التربية وعلم النفس يستحيل عدّه مقياساً ثابتاً ودقيقاً كالمقاييس الطبيعية (المتر، الكغ، ... إلخ). ففي القياس الفيزيائي يمكن الحصول على قياسات ثابتة (نسبياً) لمجموعة من الأسباب بينها أن الصفات المادية مستقرة نسبياً ويمكن إخضاعها للقياس المباشر، وبينها ما يتصل بأدوات القياس ومستوى الدقة الذي وصلت إليه. ويعدّ القياس النفسي والتربوي أقل ثباتاً من القياس الفيزيائي لأسباب عديدة بينها ما يتصل بعدم استقرار السمات أو الصفات المقيسة بمرور الوقت، وبينها ما يتصل بخطأ العينة أو المعاينة، بالإضافة إلى أخطاء

التطبيق والتصحيح وغيرها. ويمكن إجمال مصادر الخطأ في القياس النفسي والتربوي على النحو التالي:

1. أخطاء في المقياس نفسه: فالمقياس هو عينة من البنود وقد لا تكون هذه العينة ممثلة لجميع البنود وهذا ما يسمى بخطأ العينة أو المعاينة Sampling Error.
2. الأخطاء الناجمة عن شروط التطبيق: فإذا لم تراع شروط التطبيق كما تنص عليها تعليمات الاختبار أو روعيت بصورة غير كافية فقد تؤدي إلى نتائج غير صحيحة.
3. الأخطاء الناجمة عن التصحيح: وهذه الأخطاء تظهر بخاصة في الأسئلة المقالية التي تفسح مجالاً واسعاً لظهور العوامل الذاتية للمصحح واحتمال نشوء الاختلاف بين مصحح وآخر في تقدير الدرجات، أو حتى عند المصحح نفسه من وقت لآخر.
4. الأخطاء الناجمة عن المفحوص نفسه: وهذه الأخطاء كثيرة ومتنوعة منها ما يتصل بقلق الامتحان ومنها ما يتصل بالتعليمات والميل إلى التخمين والدافعية للأداء ودرجة التأثر بالعوامل الطبيعية كالحرارة والبرودة والضجيج... وغيرها. هذا بالإضافة إلى عدم ثبات واستقرار السمة المقيسة عند المفحوص التي يمكن أن تتأثر بالعديد من العوامل كالتدريب والتذكر والنسيان وغيرها.

إن واحداً أو أكثر من مصادر الخطأ السابقة يمكن أن يؤثر في جميع الطلاب بالدرجة نفسها ويطلق عليه في هذه الحالة اسم الخطأ المنتظم Systematic Error وقد يختلف تأثيره من طالب لآخر سواء من حيث المقدار أو الاتجاه (سلبي أو إيجابي) ويطلق عليه في هذه الحالة اسم الخطأ العشوائي Random Error. ومن الواضح أن بعض العوامل المتصلة بالاختبار نفسه وبشروط التطبيق والتصحيح يمكن ضبطها إلى حد كبير وتقليل الأخطاء الناجمة عنها، وأن بعض العوامل ولاسيما تلك المتصلة بالمفحوص نفسه يصعب

ضبطها والتحكم بها ويتعذّر بالتالي تحاشي احتمالات الخطأ التي يمكن أن تنتج عنها. ومهما يكن من أمر فإن التباين في الدرجات التي يعطيها المقياس يمثّل التباين الحقيقي في أداء المفحوصين كما يمثّل التباين الناتج عن الخطأ أو تباين الخطأ Error Variance. ويمكن توضيح الثبات من خلال مفهوم التباين على النحو التالي:

$$ع^2_{س} = ع^2_{سق} + ع^2_{خم} + ع^2_{خع} \quad \text{(المعادلة 5 – 2)}$$

حيث:

$ع^2_{س}$ = التباين الكلي.

$ع^2_{سق}$ = التباين الحقيقي في علامات الطلاب الناتج عن فروق حقيقية بينهم في السمة المقيمة.

$ع^2_{خم}$ = التباين الحقيقي في علامات الطلاب الناتج عن تباينهم في درجة التأثر بالعوامل المسببة للخطأ المنتظم.

$ع^2_{خع}$ = التباين في الخطأ العشوائي.

فإذا افترضنا أن التباين الناتج عن الخطأ العشوائي يساوي صفراً وهذا مجرد افتراض يستحيل تحقيقه في الواقع في القياس النفسي والتربوي، فإن التباين الكلي يساوي التباين الحقيقي، وهذا يعني أن الاختبار ثابت تماماً أو أنه يقيس بدقة تامة ودون أخطاء. ويمكن القول بلغة الإحصاء: إن الثبات هو نسبة التباين الحقيقي إلى التباين الكلي، مع ملاحظة أن التباين الذي ينطوي عليه مفهوم الثبات يشمل التباين الناتج عن فروق حقيقية في أداء المفحوصين والتباين الناتج عن الخطأ المنتظم. ومن الواضح أن الثبات يزداد كلما نقص تباين الخطأ (أي التباين الناتج عن الخطأ العشوائي) وينخفض كلما ازداد هذا الأخير. ومن

المفيد الإشارة هنا إلى أنه إذا انعدم التباين الناتج عن الخطأ العشوائي وانعدم التباين الناتج عن الخطأ المنتظم(وهذا مجرد افتراض) فإن التباين الحقيقي ينسب كلياً إلى السمة المقيسة (أي التباين الحقيقي يعادل التباين الظاهري) وفي هذه الحالة يقيس الاختبار ما وضع لقياسه أي يصدق تماماً وهذا هو المقصود بصدق الاختبار.

وبالإضافة إلى مفهوم التباين ودوره في توضيح معنى الثبات من المفيد ملاحظة النقاط التالية لإلقاء المزيد من الضوء على مفهوم الثبات:

1. يرتبط الثبات بنتائج أداة القياس والتقويم لا بالأداة ذاتها. وبدلاً من الحديث عن ثبات الأداة بحد ذاتها من الأصوب الحديث عن ثبات النتائج. وقد يكون للأداة الواحدة أكثر من معامل ثبات واحد وذلك تبعاً للمجموعة والموقف الاختباري. وبما أن لكل عينة من الأفراد خصائصها فإنه سيكون من المفيد الحصول على أكثر من معامل ثبات للاختبار الواحد لدى أكثر من عينة. فمعامل الثبات المستخرج من أداء عينة من الأطفال قد يختلف عن نظيره المستخرج من أداء عينة من الراشدين، ومعامل الثبات المستخرج من أداء عينة من الذكور، أو المعاقين قد يختلف عنه لدى الإناث أو العاديين، وهكذا...
2. تقدير ثبات الاختبار يرتبط بالاستعمال الخاص له. فلبعض الاستعمالات قد نهتم بالسؤال عن درجة ثبات نتائج الاختبار عبر فترة من الزمن، ولاستعمالات أخرى قد نهتم بالسؤال عن درجة ثبات النتائج من خلال عينات مختلفة من الأسئلة أو مقدّرين مختلفين. ولاستعمالات أخرى قد نهتم بالاتساق الداخلي للأداة ذاتها (Gronlund,1971,p.101). ولا يعني هذا أن الاختبار الواحد لا تصلح له إلا طريقة واحدة من طرق الثبات. وقد يكون من الأنسب في حالات كثيرة استخدام أكثر من طريقة واحدة للاختبار الواحد.

3. الثبات هو مفهوم إحصائي. بخلاف الصدق الذي يعتمد على التحليل المنطقي وقد يتطلب العمل الإحصائي.

العلاقة بين الصدق والثبات:

تشير المعادلة الخاصة بالتباين الكلي التي وردت سابقاً (المعادلة 5 – 2)إلى أنه إذا افترضنا انعدام التباين الحقيقي في السمة المقيسة وانعدام الصدق

(أي $ع^2_{س ق}$ = صفراً) فإن التباين في العلامات لا يعود إلى فروق حقيقية بين الأفراد في السمة المقيسة بل يعود إلى فروق ناجمة عن الخطأ. وفي هذه الحالة يمكن القول: إن الاختبار "ثابت" ولكنه غير صادق. وهذا يعني إن الاختبار الذي قد يكون " ثابتاً " أو على درجة ما من الثبات قد لا يكون صادقاً أو لا يتمتّع بأي درجة من الصدق. ويمكن من الناحية المنطقية إثبات أن الاختبار قد يكون "ثابتاً" أو يتمتع بدرجة من الثبات دون أن يكون صادقاً، فمثلاً اختبار المعلومات التاريخية الذي يقيس فهم المفردات والتراكيب اللغوية قد يعطي نتائج متسقة في حال تكراره ويكون على درجة عالية من الثبات ولكنه غير صادق في قياس المعلومات التاريخية. وإذا أعطى عقرب الساعة إشارة للوقت تزيد بمقدار 50 دقيقة لعدة مرات، أو أعطى ميزان البقال وزناً واحداً خاطئاً لعدة مرات فيمكن القول: إن النتيجة ثابتة مع أنها غير صادقة بالتأكيد.

يعتمد الصدق جزئياً على الثبات، ويمثّل الثبات شرطاً ضرورياً وغير كاف للصدق. ومن الواضح أن المقياس الذي يعطي نتائج غير متسقة لا يمكن أن يعطي نتائج صادقة على حين أن المقياس الذي يعطي نتائج صادقة لابد أن يعطي نتائج متسقة ويتمتع بصفة الثبات. وبطبيعة الحال فإن الزيادة في معامل الثبات قد تؤدي إلى الزيادة في معامل الصدق ولكنها لا تحتّم هذه الزيادة. فقد تنتج الزيادة في معامل الثبات عن ازدياد التباين الناتج عن الخطأ المنتظم كما يظهر بصورة واضحة في المعادلة السابقة الخاصة بالتباين.

طرائق الثبات:

يميّز الباحثون عادة بين طرائق عدة في حساب الثبات أهمها طريقة الإعادة Test – Retest Method، وطريقة إجراء مقياس معادل Equivalent Forms، وطريقة التنصيف Split – half Method، وطريقة الاتساق الداخلي باستخدام معامل كودر – ريتشارسون ومعامل الفا Kuder – Richardson Reliability and Coefficient Alpha، بالإضافة إلى الطريقة التي تعرف بثبات المصححين Scorer Reliability. وتتطلب الطريقة الأولى من هذه الطرائق وهي طريقة الإعادة تطبيقين للاختبار، وتتطلب الثانية وهي طريقة الشكل المعادل تطبيقاً للاختبار ولصورة معادلة له، في حين أن الطرائق الثلاث الأخرى تتطلب تطبيقاً واحداً للاختبار. وسوف نتناول فيما يلي هذه الطرائق المختلفة بالدراسة:

1. طريقة الإعادة:

وتقوم هذه الطريقة على إعادة تطبيق المقياس نفسه على جماعة واحدة من المفحوصين ثم حساب معامل الثبات وهو معامل الترابط بين العلامات التي ينتهي إليها التطبيق الأول والعلامات التي ينتهي إليها التطبيق الثاني للمقياس أو الاختبار. ويسمى معامل الثبات الذي يحسب بهذه الطريقة معامل الاستقرار Coefficient of Stability. وقد وجّهت لهذه الطريقة انتقادات عديدة منها:

1. صعوبة خفض أثر التغيرات التي قد تقع في الفترة الفاصلة بين التطبيق الأول والتطبيق الثاني للمقياس، فقد يكتسب المفحوصون أو بعضهم خلال هذه الفترة خبرات جديدة. وقد يتعرض بعضهم للنسيان ولاسيما إذا طالت الفترة، وأما إذا قصرت فإن تذكر الطلاب للأجوبة الأولى قد يؤدي دوره وهذا ما يرفع معامل الثبات بصورة مصطنعة. والواقع أنه من الصعب تعيين فترة فاصلة مثلى بين التطبيق الأول والتطبيق الثاني للاختبار وإن كان

المتبّع بصورة عامة هو أن لا تقل هذه الفترة عن أسبوع ولا تزيد عن ستة أشهر. هذا مع الإشارة إلى أن المتبع في اختبارات التحصيل هو أن لا تزيد هذه الفترة على شهر واحد ولا تقل عن بضعة أيام.

2. ثمة عوامل عديدة تؤثّر في أداء المفحوصين كما هو معلوم، فإذا أعطي الاختبار في مناسبتين مختلفتين فإن لهذه العوامل فرصتين للتأثير في النتائج، وبالتالي فإن مصادر خطأ عديدة تفعل فعلها وتؤدي نظرياً إلى مضاعفة أخطاء القياس ولاسيما تلك الأخطاء الناجمة عن شروط التطبيق والأخطاء الناجمة عن المفحوص نفسه كالقلق والدافعية والحالة الصحية وغيرها.

3. تتطلب هذه الطريقة جلستين اختباريتين وتعدّ غير اقتصادية من ناحية الوقت، وقد تتعب المفحوص. هذا بالإضافة إلى احتمال انخفاض مستوى الدافعية للأداء الاختباري في الجلسة الاختبارية الثانية.

إن إدراك العيوب التي تعاني منها طريقة الإعادة في حساب الثبات لابد أن يدفع الباحث أو العامل في القياس إلى التصدّي لها والعمل على التخفيف من وطأتها إلى الحدود القصوى المتاحة. وقد يكون من المفضل عند استخدام هذه الطريقة ألا يقتصر الباحث على حساب الثبات على مدى فترة زمنية واحدة، بل على مدى فترتين أو أكثر، كأن يحسب معامل الثبات لفترة عشرة أيام(أي بفاصل قدره عشرة أيام بين التطبيق الأول والتطبيق الثاني للاختبار)، ولفترة شهر واحد، ولفترة ثلاثة أشهر. ولمواجهة احتمال تأثر أفراد العينة بإعادة الاختبار لأكثر من مرة سواء بالاتجاه الإيجابي أم السلبي (كأن يتحسن أداؤهم نتيجة التدرب على الاختبار، أو يضعف نتيجة للملل أو اللامبالاة)، فقد يكون من المناسب استخدام عينتين متجانستين (أو أكثر) يطبق الاختبار ثم يعاد تطبيقه على كل منهما بفواصل زمنية مختلفة (كأن تكون الفترة الزمنية الفاصلة بين التطبيق الأول والثاني للاختبار على أفراد العينة الأولى عشرة أيام، وعلى أفراد العينة الثانية شهراً واحداً وهكذا...).

ويمكن القول: إن اختيار الأسلوب المناسب والفترة المناسبة لإعادة الاختبار ولحدود التجانس المطلوبة في العينة "يعتمد في الجانب الأكبر منه على تحليلنا السيكولوجي لما يقيسه الاختبار، وكيفية قياسه، ولمعلوماتنا التي يوفّرها لنا التراث عن الوظائف التي يقيسها، ونموها، وارتقائها، وتأثرها بالمتغيرات المختلفة" (فرج، 1980، ص353).

ولعل من بين الأمور التي لابد من أخذها بالحسبان عند دراسة الثبات على أية حال هو أن طريقة الإعادة تصلح لدراسة السمات التي تتسم بشيء من الثبات والاستقرار النسبي، ولاتصلح، أو لا تصلح بالدرجة نفسها، عند دراسة السمات التي تفتقر إلى الثبات والاستقرار، أو تتذبذب بفعل عوامل معينة. ومن نافلة القول أن هذه الطريقة ستكون محدودة الفائدة، وربما عديمة الفائدة، عند استخدامها في تقدير ثبات درجات مقياس للدافعية، أو للحالة المزاجية، أو لحالة القلق، أو للذكاء لدى صغار الأطفال، أو غيرها من السمات او الحالات التي يكون لعامل الزمن دوره وأثره الواضح فيها.

2. طريقة الأشكال المتعادلة:

وترتكز هذه الطريقة على تحديد درجة الاتساق في أداء المفحوصين على عينة من البنود أو المهمات وعينة أخرى من البنود أو المهمات سحبت من المجتمع الأصلي ذاته (مجتمع البنود) الذي سحبت منه العينة الأولى. فإذا كان الاختبار عينة ممثلة للبنود المحتملة في مجال معين (أو المجتمع الأصلي للبنود) فهو مقياس ثابت للمحتوى في ذلك المجال. والطريقة الأيسر لتقدير ما إذا كان يقيس عينة ممثلة للمحتوى هي بوضع شكلين متعادلين (او أكثر) للاختبار وحساب معامل الترابط بينهما، فإذا كان الترابط عالياً بينهما فهذا يدل على أن كلا الشكلين يقيس المحتوى ذاته، ولذا فهما عينتان ثابتتان لمحتوى المجال المقيس. وبطبيعة الحال فإن المقياس الثاني يجب أن يحتوي على العدد نفسه من الأسئلة وأن تتعادل الأسئلة في كلا المقياسين من حيث مستوى الصعوبة وتمثيلها

للمحتوى وأن تكون صياغتها متماثلة، كما تتماثل أو تتعادل من حيث التعليمات والزمن المعطى للاختبار.

يميّز الباحثون عادة بين طريقتين لحساب الثبات باستخدام الأشكال المتعادلة وتتفرعان عن الطريقة الأم مدار البحث. تستخدم في الطريقة الأولى من هاتين الطريقتين أشكال متعادلة "فورية" تطبق في جلسة (أو جلسات) اختبارية متعاقبة دون وجود فترات زمنية تفصل بينها (أو بوجود فواصل ضئيلة منعاً للتأثر بعوامل التعب أو الملل، وما شابه). ومن الواضح أن هذه الطريقة تستبعد أثر التباين الناجم عن عامل الزمن، وتمتاز عن طريقة الإعادة بأنها تلغي آثار الفترة الزمنية الفاصلة بين التطبيقين الأول والثاني للاختبار، التي قد تظهر في اكتساب خبرات جديدة، أو نسيان ما تمّ تعلّمه، أو تذكر الإجابات التي أعطاها المفحوص في التطبيق الأول أو غيرها. ويطلق على معامل الثبات المستخرج بهذه الطريقة اسم معامل التعادل أو التكافؤ.

أما الطريقة الثانية لحساب الثبات باستخدام الأشكال المتعادلة فيمكن تسميتها بالأشكال المتعادلة المتعاقبة لكونها تقوم على استخدام أشكال متعادلة تطبق في جلسات اختبارية متعاقبة وبوجود فواصل زمنية طويلة نسبياً. ويطلق على معامل الثبات المستخرج بهذه الطريقة اسم معامل التعادل والاستقرار. ومن الواضح أن هذه الطريقة تتيح تقدير الاتساق في أداء المفحوصين عبر الزمن، كما تتيح تقدير الاتساق في أدائهم على عينات متجانسة من البنود. وتنطوي هذه الطريقة على كل مصادر الخطأ المحتملة بحسب جرونلند، من مثل استقرار إجراءات القياس، وثبات السمات المقيسة، وتمثيل العينة، كما تعدّ الطريقة الأكثر صرامة لحساب الثبات(Gronlund ,1971,p.105).

ومن عيوب طريقة الأشكال المتعادلة في دراسة الثبات صعوبة وضع مقياس ثان معادل للمقياس الأول سواء من حيث تمثيل البنود للسمة المقيسة، ومستوى صعوبة البنود، والصياغة، والتعليمات، والوقت المعطى للمقياس وغيرها. هذا

بالإضافة إلى أن هذه الطريقة تتطلب مضاعفة الجهد والوقت والنفقات بطبيعة الحال.

3. طريقة التنصيف:

وتقوم هذه الطريقة على تقسيم الاختبار بعد تطبيقه في جلسة اختبارية واحدة إلى نصفين يفترض أنهما متكافئان، ثم حساب معامل الارتباط بين الدرجات التي حصل عليها المفحوصون على النصف الأول والدرجات التي حصلوا عليها على النصف الثاني. ويطلق على معامل الثبات المحسوب بهذه الطريقة اسم معامل ثبات التنصيف (أو الشطر النصفي) كما يطلق عليه أحياناً اسم معامل الاتساق الداخلي.

وتشبه هذه الطريقة طريقة الأشكال المتعادلة من حيث أنها تشير إلى ما إذا كانت عينة البنود ممثلة للمحتوى المقيس، كما تغفل التغيرات التي يمكن أن تحدث لدى المفحوصين بنتيجة العامل الزمني(الثبات الزمني). وتختلف عنها من حيث أنها تطبق في جلسة اختبارية واحدة وتعدّ أكثر اقتصادية من ناحية الوقت والجهد والنفقات.

والصعوبة الكبرى التي تعترض هذه الطريقة تتمثّل في تقسيم الاختبار إلى نصفين متعادلين مع مراعاة مستوى الصعوبة ودرجة التمثيل وشكل البنود وغير ذلك. والواقع أن معظم الاختبارات يصعب تقسيمها إلى نصفين متعادلين وافتراض أن أداء الطلاب على النصف الأول يعادل أداءهم على النصف الثاني بمجرد شطر الاختبار إلى نصفين حسب التسلسل الرقمي للبنود. ومن الطرائق المتبعة في عملية التنصيف جعل جميع البنود ذات الأرقام الفردية في أحد النصفين وجعل جميع البنود ذات الأرقام الزوجية في النصف الآخر. ويرى ثورانديك وهيجن أن هذا الأسلوب مناسب ما دام الاحتمال كبيراً بأن البنود التي تتشابه في الشكل والمحتوى والصعوبة تتجمع بعضها مع بعض في الاختبار

الواحد. " ففي اختبار طويل نسبياً يصل إلى (60) بنداً أو أكثر تميل التجزئة بهذه الطريقة إلى تحقيق الموازنة في عوامل من نوع شكل الفقرة وشمول محتواها ومستوى صعوبتها. وتكون الاحتمالات حسنة أن يكون نصفا الاختبار المتشكلان بهذه الطريقة متعادلين" (ثورندايك وهيجن، 1989، ص77). ومن الطرائق المتبعة في تقسيم الاختبار بالإضافة إلى هذه الطريقة، تحديد مستوى السهولة أو معامل السهولة لكل سؤال ثم توزيع الأسئلة على نصفي الاختبار تبعاً لتكافؤ مستوى السهولة مع مراعاة التشابه في المضمون. وهذه الطريقة أدقّ من الطريقة السابقة وتتطلب جهداً زائداً.

بيد أن معامل الارتباط المحسوب بطريقة التنصيف يعطينا ثبات نصف الاختبار وليس الاختبار بكامله. ومن المعلوم أنه كلما كبرت عينة السلوك المقيسة كان قياسها أكثر ثباتاً ودقة. لذا لابد من تصحيح معامل الثبات المحسوب بهذه الطريقة باستخدام معادلة سبيرمان – براون على النحو التالي:

$$ر_{11} = \frac{2ر}{1+ر}$$

حيث يشير الرمز ($ر_{11}$) إلى معامل ثبات الاختبار بكامله والرمز (ر) إلى معامل ثبات نصف الاختبار. فإذا بلغ معامل ثبات نصف الاختبار 0.60 فإنه يصبح بعد تصحيحه باستخدام المعادلة السابقة 0.75 وهو معامل ثبات الاختبار ككل.

وهذا يشير إلى أن معامل ثبات درجات الاختبار ككل زادت قيمته عن قيمة معامل ثبات درجات كل من نصفيه على حدة.

ويظهر الجدول التالي أثر استخدام معادلة سبيرمان – براون في رفع قيم معامل الثبات للاختبار الكلي حين تتراوح قيم معاملات الثبات المحسوبة لأحد نصفي الاختبار من 0.10 إلى 0.90 (الجدول رقم(10)):

الجدول رقم (10): معاملات ثبات الاختبار قبل التصحيح بمعادلة سبيرمان – براون وبعده:

قيم معامل الثبات قبل التصحيح	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
قيم معامل الثبات بعد التصحيح	0.18	0.33	0.46	0.57	0.67	0.75	0.82	0.89	0.95
الزيادة في قيم المعامل	0.08	0.13	0.16	0.17	0.17	0.15	0.12	0.09	0.05

ويتبين من قراءة الجدول السابق أن هناك ارتفاعاً في قيم معامل الثبات المصحح باستخدام معادلة سبيرمان – براون، كما يتبين أن هذا الارتفاع يكون أكبر كلما اقتربنا من القيمة الوسطى والبالغة 0.50 وأقل كلما ابتعدنا عنها نحو أحد الطرفين المتقابلين.

ومن المزايا المهمة لتطبيق معادلة سبيرمان – براون أنه يمكن عن طريقها تقدير العدد اللازم من البنود لجعل ثبات درجات الاختبار عند قيمة معينة (أي زيادة عدد البنود لرفع مستوى الثبات)، وذلك استناداً إلى قيمة معامل ثبات درجات الاختبار الأصلي (قبل التصحيح) (علام، 2000).

بيد أن معادلة سبيرمان – براون تفترض أن التباين في درجات النصف الأول (والتباين هو مربع الانحراف المعياري) مساوٍ تماماً للتباين في درجات النصف الثاني. وهذا يبقى مجرد افتراض حتى لو بدا نصفا الاختبار على درجة كبيرة من التكافؤ. ولذلك فقد يستعاض عنها بمعادلة جثمان (والتي تنسب أيضاً إلى فلاناجان) التي تستبعد هذا الافتراض. ويمكن صياغتها على النحو التالي:

$$ر_1 = 2 \left(1 - \frac{ع^2أ + ع^2ب}{ع^2ك}\right)$$

حيث يشير الرمز ($ر_1$) إلى معامل ثبات الاختبار ويشير الرمز ($ع^2أ$) إلى تباين درجات أحد نصفي الاختبار، والرمز ($ع^2ب$) إلى تباين درجات النصف الثاني، والرمز ($ع^2ك$) إلى تباين الدرجات الكلية في الاختبار.

ولإيضاح كيفية استخدام معادلة جتمان في إيجاد قيمة معامل الثبات لدرجات الاختبار ككل نفترض أننا قمنا بتقسيم الاختبار إلى نصفين يتضمّن الأول منهما الأسئلة ذات الأرقام الفردية، ويتضمّن الثاني الأسئلة ذات الأرقام الزوجية، ثم حسبنا قيمة الانحراف المعياري لدرجات كل من نصفي الاختبار، كما حسبنا قيمة الانحراف المعياري لدرجات الاختبار ككل، ثم ربّعنا هذه القيم للحصول على التباين الخاص بدرجات كل من نصفي الاختبار ودرجات الاختبار ككل فحصلنا على النتائج التالية:

ع أ للنصف الأول = 1.5

$ع^2$أ للنصف الأول = 2.25

ع ب للنصف الثاني = 1.65

$ع^2$ب للنصف الثاني = 2.72

ع ك للاختبار الكلي = 2.85

$ع^2$ك للاختبار الكلي = 8.12

وبالتعويض عن هذه القيم في المعادلة السابقة نحصل على ما يلي:

$$ر_1 = 2 \left(1 - \frac{2.72 + 2.25}{8.12}\right)$$

$$ر_1 = 2 \left(1 - \frac{4.97}{8.12}\right)$$

= 0.78 تقريباً

ويتبين من المثال السابق أن استخدام معادلة جتمان يتيح تقدير ثبات الاختبار بكامله دون الحاجة لإعادة تصحيح الطول أو استخدام معادلة تالية لها لتصحيح الطول.

ومن الطرائق المتبعة في حساب ثبات التنصيف التي تتيح بدورها تقدير ثبات الاختبار بكامله دون الحاجة لإعادة تصحيح الطول أو استخدام معادلة تالية لها لهذا التصحيح طريقة رولون. وتقوم هذه الطريقة على حساب تباين الفرق بين درجات المفحوصين في النصف الأول ودرجاتهم في النصف الثاني للاختبار، وحساب تباين درجات الاختبار ككل. والصيغة الخاصة بمعادلة رولون هي التالية:

$$ر ك = 1 - \frac{ع^2 ف}{ع^2 ك}$$

حيث يشير الرمز ر ك إلى معامل ثبات الاختبار كله.

ويشير الرمز $ع^2$ ف إلى تباين الفرق بين درجات الأفراد في النصف الأول ودرجاتهم في النصف الثاني للاختبار.

ويشير الرمز $ع^2$ ك إلى تباين الاختبار ككل.

ويمكن ملاحظة أن نص هذه المعادلة يظهر الصلة بينها وبين التعريف الخاص بتباين الخطأ الذي سبقت الإشارة إليه. فأي فرق بين درجات المفحوصين في النصفين يمثّل تباين الخطأ. وبحساب تباين الفروق بين درجات المفحوصين في النصفين وتقسيمه على التباين الكلي يمكن الحصول على نسبة تباين الخطأ في الدرجات. وحين يطرح هذا التباين في الخطأ من الواحد الصحيح يعطينا نسبة التباين الحقيقي للاختبار، والذي يعادل معامل ثبات هذا الاختبار.

فإذا كان تباين الفرق بين درجات المفحوصين في النصف الأول ودرجاتهم في النصف الثاني هو 4.28، وكان التباين الكلي لدرجات الاختبار مثلاً هو 22.88 فإن معامل ثبات الاختبار بهذه الطريقة يحسب بتطبيق المعادلة السابقة على النحو التالي:

$$\text{ر ك} = 1 - \frac{4.28}{22.88} = 0.81 \text{ تقريباً}$$

ومن الطرائق الأخرى المتبعة في تقدير ثبات الاختبار من تطبيقه لمرة واحدة فقط طريقة هورست وطريقة موزير. ولا يتسع المقام للوقوف عند هاتين الطريقتين في الكتاب الحالي، و يمكن الرجوع إليهما في الكتب والمراجع الإحصائية المتخصصة.

4. طريقة كودر – ريتشاردسون ومعامل ألفا:

وتقوم طريقة كودر – ريتشاردسون على تقسيم الاختبار الواحد إلى عدد كبير من الأجزاء (بدلاً من تقسيمه إلى جزئين أو نصفين متعادلين فقط)، وبحيث يتكون كل جزء من هذه الأجزاء، من بند واحد فقط من بنود الاختبار. وتتطلب هذه الطريقة حساب ارتباط كل من هذه الأجزاء أو البنود مع بعضها البعض، وكذلك ارتباط كل منها مع الاختبار ككل. وبذلك فإن الفكرة الموجهة لهذه الطريقة تتلخّص في دراسة التجانس أو الاتساق في إجابات

المفحوصين على سائر البنود التي يتألف منها الاختبار، وهو ما يعرف بدراسة التجانس أو الاتساق الداخلي بين بنود الاختبار Interitem Consistency. ومن الواضح أن هذه الطريقة تفترض أن الاختبار احادي البعد ويقيس سمة أو وظيفة واحدة فقط، وأن بنوده جميعها تقيس هذه السمة الواحدة.

من المعلوم أن الاتساق بين البنود يتأثر بمصدرين اثنين من مصادر تباين الخطأ وهما:

1. خطأ العينة أو المعاينة والذي يرتبط بمدى تمثيل البنود للمحتوى المقيس (وهذا ما يظهر عند استخدام طريقة الشكل المعادل وطريقة التنصيف في حساب الثبات).
2. عدم تجانس مجال السلوك الذي تتصدّى له بنود الاختبار. فكلما كان مجال السلوك متجانساً كان الاتساق بين البنود التي أعدت لقياسه عالياً، وكلما افتقر هذا المجال إلى التجانس ضعف الاتساق بين البنود التي أعدت لقياسه. ولتوضيح ذلك لنفترض ان لدينا اختبارين يتألف كل منهما من 20 بنداً، وأن الأول منهما يتضمّن بنوداً في الجبر فقط، في حين أن الثاني يتضمّن بنوداً في الجبر وأخرى في الهندسة ((10)بنود للجبر و(10) بنود للهندسة) فإن الاتساق الداخلي بين بنود الاختبار الأول سيكون أعلى منه بين بنود الاختبار الثاني على الأرجح. وقد يكون اداء احد التلاميذ في الاختبار الثاني أعلى في الجبر منه في الهندسة، في حين ان أداء الآخر في الهندسة قد يكون أعلى منه في الجبر وهكذا. وبالمثل فإننا لو افترضنا أيضاً، أننا طبقنا اختبارين يتألف كل منهما من 30 بنداً، وأن الاختبار الأول منهما يقيس القدرة على الفهم اللفظي فقط، في حين أن الثاني يقيس القدرة على الفهم اللفظي، كما يقيس القدرة العددية، والقدرة المكانية والإدراكية (بمعدل 10 بنود لكل قدرة). ولنفترض أيضاً أننا طبقنا هذين الاختبارين على عشرة أفراد وحصلوا على علامات تقع في المدى من صفر إلى 30 في كل من هذين الاختبارين فهل يمكن ان نتوقع اتساقاً في أداء

المفحوصين على الاختبار الثاني كما نتوقع اتساقاً في أدائهم على الاختبار الأول طالما أن الاختبار الثاني توزعت بنوده إلى ثلاث مجموعات، وأن هذه المجموعات الثلاث تقيس ثلاث قدرات مختلفة أو متمايزة، في حين أن الاختبار الأول ببنوده الثلاثين يختص بقدرة واحدة فقط، وهي القدرة على الفهم اللفظي كما أسلفنا؟

إن الإجابة عن السؤال السابق ستكون بالنفي حتماً. ذلك لأننا سنتوقع درجة عالية من الاتساق في أداء المفحوصين على الاختبار الأول فقط بوصفه اختباراً متجانساً يتناول خاصية أو قدرة واحدة، ولن نتوقع مثل هذا الاتساق في أداء المفحوصين ونتائجهم في الاختبار الثاني الذي صمّم أساساً لقدرات متعدّدة. في الوقت نفسه فإن درجات المفحوصين في الاختبار المتجانس تكون أقل غموضاً منها في الاختبار غير المتجانس أو الأقل تجانساً. فلو حصل اثنان من المفحوصين في الاختبار الثاني "غير المتجانس" على الدرجة نفسها (ولتكن الدرجة 20 مثلاً) فهل يمكن أن نستنتج أن أداء الأول يعادل تماماً أداء الثاني؟ والجواب هو بالطبع: لا. فقد تكون الدرجة 20 التي نالها المفحوص الأول هي حصيلة إجاباته الصحيحة عن البنود العشرة الخاصة بالفهم اللفظي، والبنود العشرة الخاصة بالقدرة المكانية والإدراكية فقط ودون أن يعطي أي إجابة صحيحة عن البنود العددية(أي 10+10=20). هذا في حين أن الدرجة 20 التي نالها المفحوص الثاني قد تكون حصيلة إجاباته الصحيحة عن البنود العشرة الخاصة بالقدرة على الفهم اللفظي، والبنود العشرة الخاصة بالقدرة المكانية والإدراكية دون أن يعطي أي إجابة صحيحة عن البنود العددية (أي 10 + 10+ صفر=20). وبالطبع فإن العلامة 20 يمكن أن تكون حصيلة جمع أي علامات ينالها المفحوص في كل من المجموعات الثلاث من البنود بدءاً بالعلامة صفر حتى العلامة 20.

وفيما يتصل بالاختبار الأول المتجانس نسبياً فإننا لو افترضنا حصول أحد المفحوصين على العلامة 20 فسوف تعني هذه العلامة على الأرجح أن هذا المفحوص أجاب عن البنود العشرين الأولى التي يتكون منها الاختبار، إذا كانت

بنود هذا الاختبار مرتبة وفق مبدأ التدرج في الصعوبة. وبالطبع هناك احتمال في أن يخفق هذا المفحوص في اثنين أو ثلاثة من البنود الأسهل الواقعة بين البند الأول والبند العشرين، ويعطي إجابات صحيحة عن اثنين أو ثلاثة من البنود الأصعب الواقعة بعد البند العشرين، غير أن هذه التذبذبات في أداء الفرد ستكون طفيفة بالمقارنة مع التذبذبات التي يمكن توقعها حين يكون الاختبار أقل تجانساً. ومن الواضح أن الدرجة 20 التي نالها الفرد في هذه الحالة يمكن مقارنتها بالدرجة 20 التي حصل عليها زميله مثلاً وتوضح أنهما متساويان في القدرة على الفهم اللفظي، كما أن هذه الدرجة قابلة للمقارنة مع أي درجة أخرى ينالها أي من المفحوصين الآخرين لكونها تشير إلى مستوى محدد من القدرة يعلو عن مستويات القدرة التي تعبر عنها درجات من نالوا أقل من الدرجة 20، ويهبط عن مستويات القدرة التي تعبر عنها الدرجات الأعلى من الدرجة 20.

لقد وضع كودر وريتشاردسون عدداً من المعادلات الخاصة بحساب الاتساق الداخلي بين بنود الاختبار. وهذه المعادلات جميعها تقوم على حساب درجة الاتساق في أداء الفرد من بند لآخر، وتتطلب تطبيق الاختبار لمرة واحدة فقط، ولا تتطلب إعادة تطبيقه، أو شطره إلى نصفين متعادلين. ومن أهم هذه المعادلات معادلة كودر – ريتشاردسون 20 التي تعرف اختصاراً بـ (ك – ر20)، ونصها كما يلي:

$$ر = \frac{ن}{ن-1} \times \frac{ع^2 - مج\, ص\, خ}{ع^2}$$

حيث يشير الرمز ر إلى معامل ثبات الاختبار.

والرمز ن إلى عدد البنود (الأسئلة).

والرمز $ع^2$ إلى تباين درجات الاختبار.

والرمز ص إلى نسبة الإجابات الصحيحة عن كل بند (معامل السهولة).

والرمز خ إلى نسبة الإجابات الخاطئة عن كل بند (معامل الصعوبة).

والمثال التالي يوضح كيفية تطبيق هذه المعادلة (هذا مع الإشارة إلى أن تطبيق هذه المعادلة قد يعطي تقديراً غير سليم للثبات إذا كان الاختبار قصيراً أو اعتمد على السرعة):

إذا طبقنا اختباراً مؤلفاً من 30 بنداً على عينة من الأفراد، وحسبنا الانحراف المعياري لدرجات هذا الاختبار فبلغ 5.92، كما حسبنا مجموع حاصل ضرب نسبة الإجابات الصحيحة × نسبة الإجابات الخاطئة لكل بند فبلغ هذا المجموع 6.64. فكم يكون معامل ثبات هذا الاختبار؟

بالتعويض عن المعادلة السابقة:

$$ر = \frac{30}{30-1} \times \frac{35.0464 - 6.64}{35.0464}$$

$$ر = \frac{30}{29} \times \frac{28.4064}{35.0462}$$

ر = 0.84 تقريباً.

ومن المعادلات الأخرى التي وضعها كودر وريتشاردسون معادلة كودر – ريتشاردسون 21 التي تعرف اختصاراً بـ (ك – ر21)، ونصّها كما يلي:

$$ر = \frac{ن}{ن-1}\left[1 - \frac{م\left(ن-م\right)}{ن \times ع^2}\right]$$

حيث يشير الرمز ر إلى معامل ثبات الاختبار.

والرمز ن إلى عدد البنود (الأسئلة).

والرمزع^2 إلى تباين درجات الاختبار.

والرمزم إلى متوسط الدرجات الكلية للاختبار.

ولا يتطلب استخدام هذه المعادلة سوى إيجاد المتوسط والانحراف المعياري للدرجات الكلية للاختبار. غير أن هذه المعادلة تقوم على افتراض وجود تساو أو تقارب في درجات صعوبة الأسئلة التي يتضمّنها الاختبار، بمعنى أن الأسئلة جميعها يجب أن تكون من مستوى واحد من حيث الصعوبة (أو السهولة)، وهذا ما يصعب تحقيقه في معظم الحالات. لذا فإن المعادلة الأكثر ملاءمة لهذه الاختبارات، والأكثر استخداماً هي معادلة كودر – ريتشاردسون 20 التي سبقت الإشارة إليها.

ويمكن من الناحية الرياضية إثبات أن معامل الثبات المحسوب بطريقة كودر – ريتشاردسون ما هو إلا متوسط معاملات التنصيف الناتجة عن التقسيمات المختلفة للاختبار. والواقع أن طريقة كودر – ريتشاردسون تتغلب على مشكلة تعدّد طرق التجزئة النصفية للاختبار التي تؤدي إلى ظهور اختلاف في القيمة التقديرية لمعامل الثبات (كأن يقسم الاختبار مثلاً إلى بنود فردية وبنود زوجية ويقسم إلى بنود النصف الأول بحسب تسلسلها، وبنود النصف الثاني بحسب تسلسلها، فيظهر اختلافاً في معامل الثبات). " ويتمّ التغلب على مشكلة تعدّد طرق التجزئة النصفية بحسب طريقة كودر – ريتشاردسون بإيجاد قيمة تقديرية لمتوسط قيم معامل الارتباط بين كل من نصفي الاختبار وذلك لجميع طرق التجزئة الممكنة دون أن نقوم بهذه التجزئة فعلاً" (علام، 2000، ص160).

ولا تصلح طريقة كودر – ريتشاردسون للاختبارات الموقوتة أي الاختبارات التي تعتمد على السرعة في الأداء، بل تصلح لاختبارات القوة. ويمكن استخدام هذه الطريقة حين تكون الإجابة على البند ثنائية وتعطى الدرجة 1 أو صفر كأن تكون الإجابة بـ "نعم أو لا"، "موافق" أو "غير موافق"، "صح" أو "خطأ".

بالإضافة إلى ذلك يمكن استخدام هذه المعادلة حين تكون بنود الاختبار متجانسة بمعنى أنها تتناول سمة أحادية البعد Unidimensional أو تقيس متغيراً واحداً. أما في الحالات التي يتصدّى فيها الاختبار الواحد لمجموعة من السمات أو القدرات، ويشتمل على اختبارات فرعية غير متجانسة فإنه لابد من حساب ثبات الاتساق الداخلي لكل من هذه الاختبارات الفرعية على حدة، وبصورة مستقلة عن الآخر. ويطلق على معامل الثبات المستخرج في هذه الحالات معامل التجانس Homogeneity Coefficient.

ومن الواضح أنه إذا استخدمت طريقة كودر – ريتشاردسون في حساب ثبات الاختبار الذي يقيس أكثر من سمة واحدة ستهبط قيمة معامل الثبات المحسوب هبوطاً شديداً.

غير أن طريقة كودر – ريتشاردسون تصلح للاختبارات ثنائية الدرجة، أي التي يعطى البند الواحد فيها الدرجة 1 أو صفر كما سبقت الإشارة، ولا تصلح للاختبارات وأدوات القياس "غير الثنائية" أي الأدوات التي يعطى البند الواحد منها الدرجة "1" أو "2" أو "3" أو أكثر، كمقاييس الاتجاهات، وبعض استخبارات الشخصية، إضافة إلى الاختبارات التحصيلية المقالية، وسلالم التقدير، وغيرها. وقد تمكن كرونباخ من وضع معادلة تصلح لهذه الأدوات "غير الثنائية" اشتقها من معادلة كودر – ريتشاردسون 20 التي سبقت الإشارة إليها، وأطلق عليها معامل ألفا. ويمثّل معامل ألفا متوسط قيم المعاملات الناتجة عن تجزئة الاختبار إلى أجزاء بطرق مختلفة، وبذلك فإنه يمثّل معامل الارتباط بين أي جزئين من أجزاء الاختبار. والمعادلة الخاصة بحسابه هي التالية:

$$\text{معامل ألفا } a = \frac{\text{ن}}{\text{ن} - 1}\left(1 - \frac{\text{م مج ع}^2\text{ ن}}{\text{ع}^2\text{ك}}\right)$$

حيث يشير الرمز ع^2 إلى تباين كل بند من بنود الاختبار.

والرمز مج ع^2 ن إلى مجموع تباينات درجات جميع البنود.

والرمز ع^2 ك إلى تباين الاختبار ككل.

والرمز ن إلى العدد الكلي لبنود الاختبار.

ولتوضيح كيفية تطبيق معامل ألفا نقدم المثال التالي:

لنفترض أن لدينا مقياساً للاتجاهات يتألف من 6 عبارات (أو بنود) وأن هذه العبارات تصحح وفق سلم ثلاثي بحيث يعطى الجواب "موافق" الدرجة 3، والجواب "محايد" الدرجة 2، والجواب "معارض" الدرجة 1. ولنفترض أيضاً أننا طبقنا هذا المقياس على ستة أفراد وأنهم حصلوا على الدرجات التي تظهر في الجدول رقم (11)، فإن الخطوات التي يجب اتباعها لحساب معامل ألفا تتلخّص في إيجاد قيمة تباين درجات كل عبارة على حدة من جهة، وتباين درجات الاختبار ككل من جهة أخرى (مع الإشارة إلى أن التباين = مربع الانحراف المعياري). وهذا ما يظهر في الجدول التالي (الجدول رقم(11)):

الجدول رقم (11): مثال توضيحي لكيفية حساب معامل ألفا لمقياس للاتجاهات مؤلف من ستة بنود:

الأفراد	العبـــــارات						
1	2	2	1	2	2	2	11
2	3	2	3	3	2	3	16
3	2	2	2	2	1	2	11
4	1	1	1	2	1	2	8
5	2	1	2	2	2	2	11
6	2	2	2	2	1	2	11
المجموع	12	10	11	13	9	13	68
م	2	1.66	1.83	1.66	1.5	1.66	11.33
ع	0.63	0.52	0.75	0.41	0.55	0.41	2.58
ع^2	0.40	0.27	0.56	0.17	0.30	0.17	6.66

وبالتعويض عن المعادلة الخاصة بحساب معامل ألفا وحساب مجموع تباينات العبارات جميعها من جهة، وتباين الاختبار ككل من جهة أخرى نحصل على ما يلي:

$$ر = \frac{6}{6-1}\left(1 - \frac{0.17+0.30+0.17+0.56+0.27+0.40}{6.66}\right)$$

$$ر = 1.2\ (1 - 0.28)$$

$$ر = 0.864$$

5. ثبات المصححين:

لقد سبقت الإشارة إلى أن هناك العديد من أدوات القياس النفسي والتربوي التي تعاني من مشكلة تدخل العوامل الذاتية في التقدير ووضع الدرجات، التي تؤدي إلى ظهور فروق بين مصحح وآخر، بل حتى عند المصحح نفسه من وقت لآخر في هذا التقدير. والواقع أن هذه الفروق في التقدير تعدّ أحد المصادر الرئيسة لتباين الخطأ في الدرجة على الاختبار، وتمثّل العيب الأكبر من عيوب الاختبارات وأدوات التقويم التقليدية. كما تمثّل عيباً لسائر الاختبارات والأدوات "غير الموضوعية" التي تعتمد على تقدير المصحح، وليس على مفتاح للتصحيح يعطي عادة الدرجة 1 للإجابة الصحيحة وصفر للإجابة الخاطئة. ومن الواضح أن اختبارات التحصيل المقالية ومقاييس القدرة الإبداعية، والكثير من الأدوات التي تستخدم في قياس الشخصية، كالأدوات الإسقاطية، والمقابلة، وسلالم التقدير وغيرها تفسح مجالاً واسعاً لتدخل العوامل الذاتية (أو الشخصية) في التقدير، وتعاني بالتالي من هبوط، وأحياناً من هبوط حاد في مستوى ثباتها.

تظهر مشكلة الذاتية في التقدير بصورة واضحة للعيان في مقاييس القدرة الإبداعية حيث يتعين على الفاحص في هذه المقاييس أن يحدّد بنفسه مدى أصالة الاستجابة التي يعطيها المفحوص ومرونته في التفكير على الرغم من وجود "معايير" للتقدير يمكنه الاستئناس بها بطبيعة الحال. وتظهر هذه المشكلة ذاتها في الأدوات الإسقاطية حتى حين تستخدم محكات معينة يمكن اعتمادها أساساً في تفسير أو تأويل استجابة المفحوص والحكم على أدائه. ذلك لأن تأويل استجابة المفحوص والحكم على ما إذا كانت تنطبق على المواصفات أو الشروط التي يتضمّنها المحكّ أمر ذاتي يعود إلى الفاحص نفسه، وليس هناك ما يمنع احتمال ظهور فروق في هذا التأويل عند اثنين من الفاحصين يستخدمان المحكّات ذاتها في تقييم استجابات المفحوصين، بل حتى عند الفاحص نفسه من وقت لآخر.

إن السؤال الذي يطرح نفسه في الحالات السابقة وغيرها من الحالات المشابهة هو التالي: ما مدى الثقة بتقديرات الفاحصين؟ وبعبارة أدق، ما مقدار تباين الخطأ الذي يعود مصدره إلى اختلاف تقديرات الفاحصين أو المصححين؟

وللإجابة عن السؤال السابق سيكون من المفيد استخدام مقياس إحصائي يكشف عن مدى الاتساق (أو الثبات) في تقديرات الفاحصين أو المصححين، وهو ما يدخل في نطاق ما يعرف بثبات المصححين Scorer Reliability. والمقياس الإحصائي الذي يمكن استخدامه في هذه الحالة هو معامل الارتباط، أو معامل ثبات المصححين، والذي يشير إلى درجة الارتباط بين التقديرات أو الدرجات التي يعطيها أحد المصححين لإجابات مجموعة من المفحوصين على أسئلة الاختبار موضع الدراسة، والدرجات التي يعطيها مصحح آخر (أو أكثر) لهذه الإجابات ذاتها.

تظهر فائدة استخدام معامل ثبات المصححين في سائر الاختبارات التي تعاني من مشكلة الذاتية في التقدير ولاسيما تلك الاختبارات التي تجري لأغراض بحثية، أو الاختبارات التي قد تعتمد نتائجها أساساً في الحكم على

الأفراد واتخاذ قرارات معينة بشأنهم. غير أن معامل ثبات المصححين لا يمثّل الوسيلة الوحيدة لتقدير ثبات المصححين، وقد يكتفى في حالات معينة بتحديد النسبة المئوية للاتفاق بين المصححين أو المقدرين أو الذين يتولوّن إجراء المقابلة الشخصية للأفراد، ولاسيما حين يكون الهدف من عملية التقدير هو الانتقاء أو تصنيف الأفراد إلى مجموعتين من مثل "مقبول" و "غير مقبول " و"ناجح" و "راسب" و "منطوي" و "منبسط" وهكذا.

طرق أخرى لدراسة الثبات:

بالإضافة للطرق السابقة في دراسة الثبات هناك طرق أخرى تجدر الإشارة إليها، لعل من أهمها طريقة التحليل العاملي Factor analysis، والطريقة التي تعتمد على تحليل التباين Analysis of variance. ومن المعلوم أن التحليل العاملي يستهدف الوصول إلى أقل عدد من العوامل التي يمكن أن تفسر الأداء على الاختبار استناداً إلى دراسة الترابطات بين البنود التي يتكون منها هذا الاختبار، والكشف عن تلك التجمعات من البنود التي يتشبّع كل منها بعامل معين دون غيره من العوامل. ومن خلال تحديد العوامل التي تتشبّع بها تجمعات معينة من البنود يمكن تصنيف هذه البنود في مجموعات متجانسة تختص كل منها بعامل أو سمة معينة. وبطبيعة الحال، فإنه في الحالات التي يسفر فيها التحليل العاملي عن وجود عامل مشترك تتشبّع به سائر البنود يمكن النظر إلى هذه البنود على أنها تؤلف مجموعة واحدة متجانسة، وإلى الاختبار ككل على أنه على درجة لا بأس بها من الثبات أو الاتساق. وبذلك يتيح التحليل العاملي تحديد مدى التجانس بين بنود الاختبار، ويمثّل طريقة لدراسة الاتساق بين البنود والذي يمثّل جوهر مسألة الثبات.

وفيما يتصل بتحليل التباين فإنه يمثّل بدوره طريقة أخرى لدراسة الثبات أو الاتساق تقوم على إرجاع التباين الكلي لدرجات الاختبار إلى ثلاثة مصادر للتباين وهي: التباين المتعلق بالأفراد المفحوصين، والتباين المتعلق بالبنود،

والتباين الخاص بالعلاقة التفاعلية بين الأفراد والبنود. ويمكن تقدير التباين المتعلق بالأفراد المفحوصين بحساب مجموع مربعات انحرافات درجات أولئك الأفراد عن المتوسط العام لدرجاتهم. أما تباين الخطأ فيقدّر عن طريق مجموع المربعات المحسوبة لتفاعل الأفراد مع البنود. وبذلك يمكن حساب معامل ثبات الاختبار باستخدام المعادلة التالية لتحليل التباين:

$$\text{معامل الثبات} = \frac{\text{التباين المتعلق بالأفراد} - \text{التباين المتعلق بتفاعل الأفراد مع البنود}}{\text{التباين المتعلق بالأفراد}}$$

ويشير معامل الثبات المحسوب إلى مدى التجانس بين بنود الاختبار، أو اتساقه الداخلي.

هذا وقد اقترح جاكسون معامل ثبات من نوع آخر وأطلق على هذا المعامل اسم معامل الحساسية. والمعادلة الخاصة بحساب معامل الحساسية بحسب جاكسون هي:

$$\text{معامل الحساسية} = \frac{\text{التباين المتعلق بالأفراد} - \text{تباين التفاعل بين الأفراد و البنود}}{\text{تباين التفاعل بين الأفراد والبنود}}$$

ويتطلب تفسير معامل الحساسية العودة إلى مستويات الدلالة الإحصائية التي تتحدّد في ضوء فرضية التوزيع الاعتدالي للدرجات.

معامل الثبات والخطأ المعياري للقياس:

يشير الثبات كما أسلفنا إلى درجة الاتساق والدقة التي يتمتع بها الاختبار في قياس ما يقيسه، أو الدرجة التي يقيس عندها الاختبار الشيء نفسه مرة بعد أخرى وبنداً بعد بند. وليس هناك حد أدنى مقبول لدرجة الثبات التي يجب أن تتوفر في الاختبار كما يعبر عنها بمعامل الثبات وإن كان المفضل بطبيعة الحال ارتفاع هذا المعامل إلى الحدود القصوى المتاحة. ويمكن القول: إن

السؤال حول معامل الثبات المطلوب يرتبط مباشرة بنوع القرارات التي ستتخذ استناداً إلى نتائج الاختبار، فكلما كان القرار هاماً تطلب ذلك ارتفاع الثبات، وعندما نكون بصدد اتخاذ قرارات نهائية وغير قابلة للمراجعة يترتب علينا البحث عن الأداة الأكثر ثباتاً إذ لا يصح اعتماداً على مقاييس ضعيفة الثبات اتخاذ قرارات بتقديم منح دراسية أو رفض متقدّمين للدراسة في الجامعة مثلاً. وعموماً "توفّر اختبارات التحصيل المقننة وكذلك اختبارات الذكاء أعلى معاملات الثبات حيث لا تقل هذه المعاملات فيها عن 0.85 على حين لا تتجاوز في بعض اختبارات الشخصية 0.50 أما الاختبارات الصفية التي يعدّها المعلمون عادة فتصل معاملات الثبات فيها في أحسن الأحوال إلى 0.60".

(Ahmann and Glock,1975,p.255)

يرتبط مفهوم معامل الثبات ارتباطاً وثيقاً بمفهوم الخطأ المعياري للقياس Standard error of measurement (SEM). فإذا أخذنا في الاعتبار أن الخطأ المعياري للقياس هو الانحراف المعياري لأخطاء القياس، وأن معامل الثبات هو ذلك الجزء من التباين الكلي الذي لا ينتج عن أخطاء القياس أمكن القول: إنه كلما تناقص التباين الناتج عن الخطأ، أي انخفض الخطأ المعياري للقياس، ارتفع معامل الثبات، وكلما ازداد تباين الخطأ بالنسبة للتباين الكلي انخفض معامل الثبات. ويمكن استخدام معامل الثبات في تقدير الخطأ المعياري للقياس ومن ثم تحديد مدى دقة كل علامة على حدة طالما أن الخطأ المعياري للقياس يشير إلى زمرة الخطأ المحيطة بكل علامة. وهذا يعني أن الخطأ المعياري للقياس يشير إلى المدى المحتمل للخطأ في علامة الفرد، وبالتالي فهو يفيد في تقدير العلامة الحقيقية للفرد. ويحسب الخطأ المعياري للقياس بالمعادلة التالية:

$$ع_خ = ع \times \sqrt{1 - ر_ث}$$

حيث يشير الرمز ع خ إلى الخطأ المعياري للقياس.

والرمز ع إلى الانحراف المعياري للدرجات.

والرمز ر ث إلى معامل ثبات الاختبار.

أي أن الخطأ المعياري للقياس = حاصل ضرب الانحراف المعياري للدرجات في الجذر التربيعي لمعامل الثبات مطروحاً من الواحد الصحيح. فإذا كان الانحراف المعياري للدرجات هو 5، وكان معامل ثبات الاختبار هو 0.91 فإن الخطأ المعياري للقياس سيكون:

بالتعويض عن المعادلة السابقة:

$$\text{ع خ} = 5 \times \sqrt{1 - 0.91}$$

$$\text{ع خ} = 5 \times \sqrt{0.09}$$

$$\text{ع خ} = 5 \times 0.3 = 1.5$$

ويمكننا القول في هذه الحالة إن الدرجة الحقيقية للفرد لا تختلف عن الدرجة التي حصل عليها إلا بمقدار درجة ونصف زيادة أو نقصاناً، وذلك في 68٪ من الحالات أو بنسبة 2: 1 من المرات. ذلك لأن 68٪ من المساحة تحت المنحني الاعتدالي تنحصر بين + 1ع و – 1 ع، أي بين انحراف معياري واحد (أو وحدة خطأ معياري واحد) إلى يمين المتوسط وانحراف معياري واحد (أو وحدة خطأ معياري واحد) إلى يسار المتوسط.

فلو كانت الدرجة التي حصل عليها الفرد هي 50.

فإن درجته الحقيقية في 68٪ من الحالات ستتراوح بين 50 +1.5 و50 – 1.5.

وهذا يعني أننا على ثقة بنسبة 68٪ بأن الدرجة الحقيقية لهذا الفرد تقع في المدى من 51.5 إلى 48.5.

ولو أردنا حساب درجة الفرد الحقيقية بدرجة أعلى من الثقة، كأن تكون بنسبة 95٪ من الثقة مثلاً، فإنه سيكون بإمكاننا أن نفعل ذلك استناداً إلى خواص المنحنى الاعتدالي حيث تنحصر 95٪ تقريباً من المساحة تحت المنحني الاعتدالي بين +2ع و – 2ع، (أي بين وحدتين للخطأ المعياري فوق المتوسط ووحدتين تحته). وفي هذه الحالة سنضرب قيمة الخطأ المعياري للقياس بـ2ع بدلاً من 1ع، أي 1.5×2 = 3 وسوف تتراوح الدرجة الحقيقية لهذا الفرد بين 50 + (1.5×2) و 50 – (1.5×2) أي بين 53 و 47.وسيمكننا القول: إننا على ثقة بنسبة 95٪ (أو في 95٪ من الحالات) بأن الدرجة الحقيقية لهذا الفرد لا تختلف عن الدرجة التي حصل عليها إلا بمقدار 3 درجات زيادة أو نقصاناً، أي أنها تقع في المدى من 53 إلى 47.

ويتبين مما سبق أن الخطأ المعياري للقياس يشير إلى مقدار التذبذب (أو الخطأ) المحتمل في درجة الفرد إذا ما أعيد تطبيق الاختبار عليه عدداً كبيراً من المرات. في الوقت نفسه، فإن المدى الذي تقع فيه الدرجة الحقيقية للفرد يزداد بزيادة درجة الثقة المطلوبة. فلو أردنا زيادة درجة الثقة لتصل إلى ما يقارب 100٪ من الحالات (أي في المدى من +3ع إلى – 3ع حيث تنحصر 99.7٪ تقريباً من مساحة المنحنى الاعتدالي)، فإن المدى الذي ستقع فيه درجة الفرد في مثالنا سيتراوح من 50+ (3×1.5) إلى 50 – (3×1.5) أي من 54.5 إلى 45.5. هذا مع الإشارة إلى أنه يكتفى عادة بالمدى من +1ع إلى – 1ع، وفي حالات قليلة فقط يفضل توسيع هذا المدى كأن نرغب في مقارنة الدرجات التي يحصل عليها الفرد الواحد في اختبارات متعدّدة من خلال رسم الصفحة النفسية له، أو غيرها من الحالات.

ومن المفيد الإشارة إلى أن استخدام الخطأ المعياري للقياس والتعبير عن أداء الفرد بصورة زمرة من الدرجات تقع ضمنها درجته الحقيقية يقف في وجه التشديد المفرط على درجة رقمية واحدة تعبر عن أداء الفرد وما قد ينطوي عليه من مدلولات. ولعل مما يؤكد ذلك تزايد عدد الاختبارات المنشورة حالياً التي توفّر ثبتاً للدرجات، لا بصورة أرقام مفردة، بل بصورة زمر من الدرجات التي يمكن أن تقع ضمنها الدرجة الحقيقية للفرد. وقد يشير هذا الأمر بالذات إلى أحد ملامح التطوّر الذي تشهده حركة القياس في الظرف الراهن.

العوامل المؤثرة في الثبات:

ثمة عوامل عديدة تؤثّر في ثبات المقياس منها:

1. **عدد البنود:**

فكلما ازداد عدد البنود ارتفع معامل الثبات. وهذا يعود إلى أن الاختبار الأطول يقدّم عينة أكثر تمثيلاً للسلوك المقيس ويعكس الفروق الحقيقية في هذا السلوك بصورة أفضل، وتكون الدرجات المتحصّلة على هذا الاختبار أقل تأثراً بالتخمين.

2. **انتشار الدرجات:**

كلما ازداد انتشار الدرجات ارتفعت قيمة معامل الثبات وتضاءل الأثر الذي تحدثه أخطاء القياس على مركز الفرد داخل المجموعة. ومن الواضح أن كل ما من شأنه أن يضعف احتمال تذبذب مراكز الأفراد داخل المجموعة من قياس لآخر يرفع الثبات فإذا ازداد انتشار الدرجات تضاءل هذا الاحتمال. لاحظ المثال التالي لمجموعتين من الدرجات حصلت عليها مجموعتان من الأفراد:

88	88
81	87
72	85
68	82
56	81
48	79
41	78
33	77

فإذا أعدنا القياس للمرة الثانية يزداد احتمال تغير مراكز الأفراد في المجموعة الأولى أكثر من الثانية وبالتالي يضعف الثبات فيما يتصل بنتائج المجموعة الأولى.

3. صعوبة البنود:

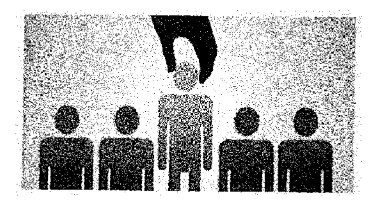
فالاختبارات الصعبة جداً أو السهلة جداً تعطي درجات ضعيفة في ثباتها لأنها تؤدي إلى انتشار محدود للدرجات وتظهر فروقاً طفيفة بين الأفراد. وعموماً إذا شاء واضع المقياس أن يزيد من ثباته عليه أن يضع أسئلة يتراوح مدى صعوبتها بين 0.30 و0.70 وبحيث يقترب معامل صعوبة الاختبار ككل من 0.50.

4. الموضوعية في التصحيح:

فلكي يعطي المقياس نتائج ثابتة يجب ألا تختلف الدرجات من مصحح لآخر أو من وقت لآخر. ومن المعلوم أن الاختبارات الموضوعية عموماً تحقق قدراً عالياً من الثبات بخلاف الاختبارات المقالية التي تفسح المجال لتدخل العوامل الذاتية للمصحح.

الفصل السابع

معايير الاختبارات النفسية



الفصل السابع
معايير الاختبارات النفسية

تحتل المعايير في القياسات النفسية والتربوية مكانة بالغة الأهمية نظراً لأنها تتيح تفسير درجة الفرد في الاختبار الذي استخرجت معاييره وتعرّف مركزه النسبي بين زملائه (أو نظرائه) في المجموعة الصفية أو العمرية التي ينتمي إليها. وتمثّل المعايير من هذه الناحية الأسس التي يمكن اعتمادها في الحكم على أداء الأفراد التي تُكسب الدرجات المتحصّلة على الاختبار معناها ودلالتها. ومن المؤكد أن درجات الأفراد في اختبار معين التي درجت العادة على تسميتها بالدرجات الخام لا تنطوي بحد ذاتها على أي معنى أو دلالة، وأنّ مجرد نسبتها إلى النهاية العظمى لدرجات الاختبار لا تضمن إعطاء صورة صادقة عن مستويات الأفراد، بل يمكن أن تعطي صورة مضللّة عن تلك المستويات في حالات معينة. والواقع أن المعايير التي تزوّد بها الاختبارات النفسية والتربوية المقننة هي التي تجعل من هذه الاختبارات أدوات قياس علمية بمعنى الكلمة و"موازين" قادرة على تحديد مستويات الأداء بدرجة عالية من الدقة. ومن هذه الناحية فإن الدور الذي تؤديه المعايير في القياسات النفسية والتربوية قد لا يقل أهمية عن الدور الذي تؤديه المعايير في القياسات الفيزيائية المختلفة كمعايير الأوزان والأحجام والأبعاد والزمن وغيرها. إذ لولا هذه المعايير وما توفّره من وحدات قياس متساوية ومنتظمة لانعدمت الركيزة التي تستند إليها عملية القياس برمتها وبطلت هذه العملية من أساسها.

ونظراً للأهمية الفائقة التي تنطوي عليها المعايير في الاختبارات النفسية والتربوية عامة – ولا يستثنى منها تلك الاختبارات التي تتصدّى للشخصية الكلية أو لجوانب معينة منها – فإننا سنسعى في هذا الفصل إلى إبراز أهميتها ودراسة الأنواع المهمة منها، كما سنسعى إلى إلقاء بعض الضوء على طرائق

استخراجها، والشروط الواجب توفرها لضمان صدقها، وعدم إساءة تفسيرها، واستخدامها على النحو السليم.

معنى المعايير وأهميتها:

للتعرّف على معنى المعايير وأهميتها في القياسات النفسية والتربوية لابد من العودة إلى الدرجة الخام التي يحصل عليها المفحوص في اختبار معين التي تدل على نتيجته النهائية في هذا الاختبار. وتحسب الدرجة الخام عادة بالجمع البسيط لعدد الإجابات الصحيحة للمفحوص عن بنود (أو أسئلة) الاختبار إذا كان الاختبار من نوع اختبارات الأداء الأقصى (أو اختبارات القوة) من مثل اختبارات الذكاء (أو القدرة العقلية العامة)، واختبارات القدرات الخاصة، واختبارات التحصيل بأنواعها. أما إذا كان الاختبار من نوع اختبارات الأداء النمطي أو العادي كما هو الحال في اختبارات الشخصية الكلية، ومقاييس الاتجاهات والقيم والميول، فإن الدرجة الخام للمفحوص تحسب بالجمع البسيط لعدد الإجابات التي تصبّ في "الاتجاه" أو المنحى ذاته الذي يأخذه الاختبار على الأغلب.

غير أن الدرجات الخام التي يحصل عليها المفحوصون سواء في اختبارات النوع الأول أم اختبارات النوع الثاني لا تعبر عن وحدات قياس متساوية ومنتظمة، والفروق بينها لا تعكس بالضرورة وجود مسافات حقيقية بين الأفراد في الخاصية المقيسة. فلو حصل أحدهم في اختبار للذكاء على الدرجة (75) وحصل ثانٍ على الدرجة (100) فإن هذا لا يعني أن القدرة العقلية للأول تعادل ثلاثة أرباع القدرة العقلية للثاني. ولو حصل أحدهم في اختبار للغة الإنكليزية على الدرجة (40) وحصل ثانٍ على الدرجة (30) وثالث على الدرجة (10) فإن هذا لا يعني أن تحصيل الأول في اللغة الإنكليزية يعادل تحصيل الثاني والثالث معاً. ولو حصل رابع في هذا الاختبار نفسه على الدرجة (صفر) فإن هذا لا يعني أن معرفته معدومة تماماً في اللغة الإنكليزية، ذلك أن الصفر في القياسات النفسية

والتربوية ليس صفراً حقيقياً يدلّ على انعدام الخاصية (أو السمة) المقيسة بل هو صفر افتراضي أو عرفي يدل على عجز المفحوص عن الإجابة عن أي من أسئلة الاختبار الذي أخضع له. ومن الواضح أن الدرجة الخام للفرد لا تحمل اي معنى إذا أخذت في معزل عن أداء المجموعة العمرية أو الصفية التي ينتمي إليها ويتعذّر عن طريقها تعرّف المركز النسبي للفرد بين أقرانه في تلك المجموعة. ومجرد نسبة الدرجة الخام إلى النهاية العظمى لدرجات الاختبار — وهو ما درجت عليه العادة في الامتحانات المدرسية — قد لا تعطي إلا مؤشراً " غائماً " أو صورة " مشوّشة" لمستوى الأداء الذي تمثّله هذه الدرجة. فالدرجة الخام في اختبارات القوة مثلاً تتأثر بشدة بمستوى سهولة الاختبار وصعوبته. فإذا كانت الدرجة الخام التي حصل عليها أحدهم في أحد هذه الاختبارات هي (45) درجة وقوبلت هذه الدرجة بالنهاية العظمى لدرجات الاختبار والبالغة (60) درجة مثلاً، فإن هذه الدرجة الخام التي تبدو في الظاهر أنها مرتفعة لا تعبّر في حقيقة الأمر عن مستوى رفيع ولا تدل على تفوق صاحبها إذا كان الاختبار سهلاً نسبياً وكان متوسط الدرجات المتحصّلة على هذا الاختبار هو (48) درجة مثلاً (أي80٪ من النهاية العظمى للدرجات). كما أن الدرجة الخام (20) إذا نسبت إلى النهاية العظمى لدرجات أحد الاختبارات والبالغة (60) درجة مثلاً لا تدل على ضعف أو تدنٍ في مستوى المفحوص الذي حصل عليها إذا كان متوسط الدرجات على هذا الاختبار هو (16) درجة. بل إنها على العكس من ذلك تدل على أن هذا المفحوص يشغل مركزاً أعلى من المتوسط بين أقرانه بصورة مؤكدة.

إن الدرجة الخام التي يحصل عليها المفحوص في اختبار معين التي قد تكون مضلّلة و"عمياء" بحد ذاتها لابد من نسبتها إلى إطار عام ومقابلتها بمعيار معين يكسبها معناها ودلالتها. ويعدّ المعيار من هذه الناحية أساساً يمكن اعتماده في الحكم على أداء المفحوص من خلال مقابلة هذا الأداء كما تعبر عنه الدرجة الخام للمفحوص بأداء بقية أفراد المجموعة التي ينتمي إليها كما يتمثّل بالمتوسط الحسابي لدرجات أفراد هذه المجموعة. والواقع أن مستوى المتوسط في

المجموعة والذي يعتمد عادة أساساً للحكم على أفراد المجموعة كافة هو المعيار الأكثر شيوعاً وانتشاراً في القياسات النفسية والتربوية ويطلق عليه عادة معيار الأداء المتوسط. هذا مع الإشارة إلى أن المعايير لا تستخرج بالاعتماد على المتوسط الحسابي للدرجات فقط، بل كثيراً ما تتطلب حساب الانحراف المعياري للدرجات والذي يعدّ بدوره شكلاً من أشكال المتوسط ويعبر عن متوسط انحرافات الدرجات عن متوسطها. ولا يشير المعيار بذلك إلى الأداء المثالي للأفراد أو "الأفضل" بالنسبة لهم بل يشير إلى الأداء المتوسط أو الطبيعي أو السوي. غير أن هذا لا يعني أن المعايير تقتصر على أن تكون تعبيراً عن المتوسطات فقط، بل "هي تشير أيضاً إلى التكرار النسبي لدرجات الانحراف المتنوعة التي تقع فوق المتوسط وتحته". (Anastasi,1982,p.25). وقوائم (أو جداول) المعايير التي تزوّد بها الاختبارات المقننة عادة تستهدف تحديد مستويات الأداء بصورة شاملة ودقيقة وتعبر عن هذه المستويات برمتها من أدناها إلى أعلاها. وتتيح قوائم أو جداول المعايير تفسير الدرجات الخام التي يحصل عليها المفحوصون كافة من خلال المقابلات المعيارية، أو القيم المعيارية المقابلة، لتلك الدرجات الخام وبالتالي تقدير مختلف درجات التفوق والتخلف دون استثناء. ومن هنا يمكن القول: إن المعايير ما هي إلا قيم كمية رقمية تنطلق من مستوى المتوسط في المجموعة وتحدّد استناداً إليه مستويات الأداء المتنوعة على الاختبار بدءاً بأدناها وانتهاء بأعلاها، كما تمثّل تلك المستويات كلها.

لقد أشرنا في الفصل الخاص بتصميم الاختبار وتقنينه إلى أن عملية استخراج معايير الاختبار أو عملية التعيير (أو المعايرة أو التقنين) Standardization، وهي المرحلة الأخيرة من مراحل إعداده وتطويره، تتطلب تطبيقه على عينة واسعة وممثلة للمجتمع الأصلي أو الجماعة الأصلية التي وضع من أجلها. ويطلق على العينة التي يطبق عليها الاختبار بهدف استخراج معاييره اسم عينة التعيير او عينة التقنين:

Normative or Standardization Sample

ويمكن تعريف عينة التقنين على أنها العينة التي تؤخذ نتائج أفرادها لتكون مصدراً لمعايير الاختبار. ولا بد لاستخراج معايير الاختبار من تحويل الدرجات الخام لأفراد عينة التقنين إلى مقابلاتها المعيارية بعد إخضاعها للمعالجة الإحصائية اللازمة ودراسة تمركز هذه الدرجات وتشتّتها أو انحرافها عن المتوسط. واستناداً إلى المعايير المستخرجة يمكن تفسير معنى الدرجة الخام التي يحصل عليها كل من يخضع لهذا الاختبار لاحقاً بعد الانتهاء من عملية تقنينه. ومن نافلة القول: إن عينة التقنين لابد من اختيارها بعناية فائقة لضمان تمثيلها للمجتمع الأصلي الذي اشتقت منه بقطاعاته وشرائحه المختلفة. فإذا ضعف تمثيل عينة التقنين (أو التعيير) للمجتمع الأصلي الذي اشتقت منه فقدت المعايير المستخرجة منها الكثير من مصداقيتها وتضاءلت إمكانات الاستفادة منها في تفسير نتائج المفحوصين.

تظهر أهمية المعايير في القياسات النفسية والتربوية في النقاط الأربع التالية:

1. إنها تتيح مقارنة الفرد بغيره وتعرّف مركزه النسبي بين أقرانه في المجموعة التي ينتمي إليها سواء على مستوى المنطقة التعليمية (أو المحافظة) أم على مستوى البلد بكامله. ويتمّ ذلك عن طريق مقابلة الدرجة الخام لهذا الفرد بما يعادلها في جدول توزيع الدرجات الخام لأفراد عينة التقنين ومقابلاتها المعيارية. وأبسط مثال على ذلك هو التالي: لو كان متوسط الدرجات على اختبار للقدرة اللفظية هو (48) درجة فإن من يحصل على هذه الدرجة يكون في مستوى المتوسط، ومن يحصل على درجة أعلى منها يكون في مستوى أعلى من المتوسط، ومن يحصل على درجة أدنى منها يكون في مستوى أدنى من المتوسط.
2. إنها تتيح مقارنة الفرد بنفسه في النواحي المختلفة التي يتصدّى لها الاختبار الواحد أو مجموعة الاختبارات التي يمكن أن يخضع لها. فإذا حصل أحدهم في اختبار للقدرة العددية على الدرجة الخام (30) وفي اختبار ثان للقدرة اللفظية على الدرجة الخام (60) وكان متوسط درجات أفراد

عينة التقنين في الاختبار الأول (25) ومتوسط درجات أفراد عينة التقنين في الاختبار الثاني (65)، فإن هذا الشخص سيكون من المتفوقين في الاختبار الأول حيث تزيد درجته بمقدار (5) درجات على المتوسط، كما سيكون من المقصّرين في الاختبار الثاني حيث تنقص درجته بمقدار (5) درجات عن المتوسط هذا على الرغم من أن درجته الخام في الاختبار الأول تعادل نصف درجته الخام في الاختبار الثاني، ويمكن لهاتين الدرجتين بالتالي أن تعطيا صورة مضلّلة لهذا المفحوص فيما لو أخذتا في معزل عن معايير الاختبارين المشار إليهما.

3. إنها تساعد المعلم على تتبّع نمو التلميذ الفرد وتقدّمه في الدراسة في المجالات المختلفة من خلال مقارنة أداء هذا التلميذ بالمعايير العامة من وقت لآخر. فبالاستناد إلى المعايير العامة وبدلالة هذه المعايير يمكن على سبيل المثال معرفة العمر القرائي أو الحسابي للتلميذ عند نقطة زمنية ما (بداية فصل دراسي أو عام دراسي مثلاً) وبيان مدى تقدّمه في القراءة أو الحساب بعد فترة معينة (نهاية فصل دراسي أو عام دراسي أو أكثر).
4. إنها تتيح مقارنة مجموعة من الأفراد مع مجموعة أو مجموعات أخرى من الأفراد في جانب أو أكثر من جوانب القدرة أو الشخصية. وتتمّ هذه المقارنة بطبيعة الحال باستخدام متوسط الأداء لدى كل مجموعة ومقابلته بمتوسط أو متوسطات الأداء لدى المجموعات الأخرى.

وفي ضوء ما سبق يمكن القول: إن المعايير تمدّنا بأساس موحد للمقارنة بين الأفراد في الخاصية (أو مظهر السلوك) التي يتصدّى لها الاختبار الواحد أو مجموعة من الاختبارات، كما تمدّنا بأساس موحد للمقارنة بين الخصائص أو مظاهر السلوك التي يتصدّى لها اختبار أو مجموعة اختبارات يمكن أن يخضع له (أو لها) الفرد الواحد وتعرّف جوانب الضعف والقوة في أدائه، بل وتتبّع نموه وتقدّمه في الدراسة من وقت لآخر. هذا بالإضافة إلى أنها توفّر أساساً موحداً للمقارنة بين المجموعات المختلفة كما أسلفنا. ولعل هذا الأمر بالذات هو مما

يجنّب العامل في القياس النفسي والتربوي الكثير من احتمالات الخطأ، كما يجنّبه خطر الانزلاق نحو الكثير من الأحكام والاستنتاجات التعسّفية عند تقدير أداء الأفراد والجماعات في مجال أو أكثر أو تقدير أداء الفرد الواحد في المجالات المختلفة.

المعايير وعينة التقنين:

من الأمور التي لابد من أخذها بالحسبان عند إعداد معايير الاختبارات النفسية والتربوية واستخدامها الطبيعة النسبية للمعايير التي تعود في جانب مهم منها لخضوعها المباشر لعملية التقنين (أو التعيير) التي تتمّ على عينة من الأفراد وحساسيتها الفائقة لأي تغيير أو تعديل للشروط المحيطة بهذه العملية. والواقع أن معايير الاختبارات النفسية والتربوية تتحدّد بصورة كلية بأداء أفراد عينة التقنين بوصفها العينة التي تستخرج منها معايير الاختبار، أو إنَّها بتعبير آخر "الوليد" المباشر لهذه العينة والصورة المعبّرة عن أداء أفرادها ونتائجهم في الاختبار. ومن الطبيعي تبعاً لذلك أن يتحدّد صدق المعايير بمدى صدق عينة التقنين ودرجة تمثيلها للمجتمع الأصلي الذي سحبت منه. وكلما كبرت العينة وارتفعت درجة تمثيلها للمجتمع الأصلي بطبقاته وفئاته المختلفة ارتفعت درجة الثقة بالمعايير المستخرجة من أداء أفرادها وأمكن اعتمادها أساساً تُقارَن به نتائج المفحوصين لاحقاً بهذا الاختبار. وكلما صغرت العينة وانخفضت درجة تمثيلها للمجتمع الأصلي بقطاعاته المختلفة ضعفت درجة الثقة بالمعايير المستخرجة منها وتعذّر اعتمادها أساساً لتفسير نتائج المفحوصين والحكم على أدائهم.

وعلى هذا فإن عينة التقنين لايصح في أي حال من الأحوال أن تكون ضئيلة في عدد أفرادها، ولابد أن يصل هذا العدد إلى الدرجة التي تسمح بالقول: إنه يمثّل المجتمع الأصلي تمثيلاً واسعاً إلى الدرجة التي تتيح الحصول على نتائج متماثلة أو متقاربة إلى حد بعيد فيما لو أعيدت عملية القياس، كما تتيح

الحصول على توزيع اعتدالي للدرجات أو أقرب ما يكون إليه، ولن تؤدي زيادة عدد أفراد العينة إلى رفع مستوى تمثيلها لهذا المجتمع إلا لدرجة محدودة للغاية وبما لا يؤدي إلى تغيير النتائج المتحصّلة او يؤثر فيها تأثيراً ذا دلالة. هذا مع الإشارة إلى أنه ليس هناك قاعدة ثابتة تفيد في تحديد عدد أفراد عينة التقنين ونسبتهم إلى المجتمع الأصلي وتضمن الحصول على توزيع اعتدالي أو شبه اعتدالي للدرجات، وإنّ زيادة العدد بحد ذاتها لا تضمن دائماً تمثيلاً كافياً. فبالإضافة إلى العدد ونسبته إلى المجتمع الأصلي لابد من مراعاة شروط أخرى عديدة تضمن الوصول بالعينة المختارة إلى الدرجة القصوى أو المستوى الأعلى من مستويات التمثيل. وبين الأمور التي لابد من أخذها بالحسبان في هذا السياق جملة العوامل والمتغيرات التصنيفية التي يمكن أن تؤثر في أداء أفراد العينة المختارة مع مراعاة الحجم الحقيقي لكل منها. ومن هذه العوامل والمتغيرات: العمر، والجنس، والمستوى الاقتصادي والاجتماعي، ومستوى التعليم، والموقع الجغرافي، والانتماءات القومية، والعرقية، وغيرها.

وعموماً فإن عينة التقنين لابد أن تكون صورة مصغرة للمجتمع الأصلي بطبقاته وفئاته المختلفة مع مراعاة نسبة وجود كل منها في هذا المجتمع. وكلما اقتربت هذه الصورة من الأصل وأصبحت أكثر انطباقاً عليه أو صورة "طبق الأصل" عنه، إذا صح التعبير، ازدادت المعايير المستخرجة دقّةً وتعزّز دورها في إصدار أحكام صائبة على أداء الأفراد.

وبالإضافة إلى الدور الهائل الذي تؤديه عينة التقنين من حيث عددها ودرجة تمثيلها للمجتمع الأصلي في المعايير المستخرجة لابد من ملاحظة أن هذه المعايير تتأثر بقوة بالظروف الواقعية التي يعمل ضمنها أفراد هذه العينة. ولابد بالتالي من "توحيد" هذه الظروف أو تقنينها ومنع أي من العوامل الجانبية أو الدخيلة من التأثير في أداء المفحوصين، كما لابد من خلق ظروف مشابهة للظروف التي تمت فيها عملية التقنين إلى أقصى درجة ممكنة عند تطبيق الاختبار بعد الانتهاء من عملية تقنينه واستخدام معاييره في تفسير نتائج

المفحوصين. فإذا اختلفت هذه الظروف بدرجة ما في الحالتين انعكس هذا الاختلاف سلباً على عملية تفسير نتائج الأفراد استناداً إلى المعايير المتوافرة.

ولابد من الإشارة إلى أن البيانات الإحصائية التي توفّرها تجربة التقنين، وهي التجربة النهائية التي تجري على أفراد عينة التقنين، يتمّ إخضاعها للتحليل وتستخرج منها عادة بعض المؤشرات الإحصائية بالغة الأهمية من مثل مؤشرات الصدق والثبات بأشكالهما المختلفة التي يزوّد بها عادة دليل الاختبار المقنن بالإضافة إلى جداول أو قوائم المعايير الملحقة به. فإذا أضفنا إلى ذلك أن اختيار البنود التي ستؤلف الصورة النهائية للاختبار المقنن أو ما يعرف بـ "غربلة" البنود يتمّ عادة في ضوء البيانات الإحصائية التي توفّرها التجارب الاستطلاعية والتمهيدية السابقة لعملية التقنين حيث تخضع هذه البيانات أيضاً للتحليل وتستخرج منها مؤشرات السهولة والصعوبة، ومؤشرات القدرة التمييزية والاتساق الداخلي وغيرها، برزت أمامنا أهمية الدور الذي تؤديه العينات المختلفة التي يطبق عليها الاختبار في المراحل المختلفة لعملية تقنينه، ولاسيما الأخيرة منها، بصورة واضحة وجلية.

أنواع المعايير:

يميّز علماء القياس النفسي والتربوي بين أربعة أنواع أو أشكال رئيسة من المعايير وهي: معايير العمر Age Norms ومعايير الصف Grade Norms التي تنتمي إلى فئة المعايير الطولية أي تمتد باتجاه طولي زمني، والمعايير المئينية Percentile Norms ومعايير الدرجة المعيارية Standard Score Norms التي تنتمي إلى فئة المعايير المستعرضة أو الأفقية وترتكز إلى أداء فئة عمرية أو صفية واحدة دون الحاجة للامتداد إلى الأعمار الزمنية أو الصفوف الدراسية السابقة واللاحقة.

ولا تشير الأنواع المختلفة من المعايير إلى اختلافات جوهرية فيما بينها، وكل ما في الأمر هو أنها تتبّع طرائق وأساليب متنوعة في التعبير عن "شيء واحد " يسعى كل منها إلى تمثيله واستخراج دلالته. وسيكون من المفيد الوقوف عند كل من أنواع المعايير على حدة.

1) معايير العمر:

تظهر أهمية معايير العمر أو السن عند قياس السمات التي تنمو مع العمر على وجه التحديد من مثل الطول أو الوزن أو الذكاء.... إلخ. ويعبر معيار العمر عن القيمة الكمية المتوسطة للسمة المقيسة لدى أبناء العمر الواحد. فلو أخذنا عينة ممثلة للأطفال من عمر الرابعة مثلاً وقسنا أطوالهم ثم حسبنا المتوسط الحسابي لتلك الأطوال لكان هذا المتوسط هو معيار الطول لهذا العمر. ويمثّل مفهوم العمر العقلي الذي ارتكز إليه مقياس بينيه الأصلي للذكاء والذي يشير إلى متوسط الأداء العقلي لأبناء المجموعة العمرية الواحدة أحد أهم المعايير العمرية الشائعة في القياسات النفسية والتربوية. ومن الواضح أن العمر الزمني يتطابق مع العمر العقلي بالنسبة للشخص العادي أو المتوسط، وأنّ العمر الزمني يزيد على العمر العقلي بالنسبة للشخص الذي ينخفض أداؤه عن المتوسط في حين أن العمر الزمني يكون أقل من العمر العقلي بالنسبة للشخص المتفوق. وقد كان معيار العمر العقلي الأساس في ظهور معيار آخر بالغ الأهمية وهو معيار حاصل الذكاء أو نسبته، والذي يحسب بقسمة العمر العقلي على العمر الزمني وضرب الناتج بمئة تفادياً للكسور العشرية.

هذا ويقدّم معيار العمر التحصيلي الذي يشير إلى متوسط الإنجاز أو التحصيل كما يقيسه اختبار معين أو مجموعة من الاختبارات التحصيلية لدى مجموعة عمرية معينة مثالاً آخر للمعايير العمرية. وقد كان معيار العمر التحصيلي الأساس في ظهور معيار نسبة التعليم (أو حاصل التعليم) الذي

يستهدف الكشف عن مدى تناسب تحصيل التلميذ مع عمره الزمني، ويحسب بالطريقة نفسها التي يحسب بها حاصل الذكاء وذلك على النحو التالي:

$$\text{نسبة التعليم (أو حاصل التعليم)} = \frac{\text{العمر التحصيلي}}{\text{العمر الزمني}} \times 100$$

كما كان معيار العمر التحصيلي الأساس في ظهور معيار آخر وهو حاصل التحصيل أو نسبته والذي يستهدف الكشف عن مدى تناسب تحصيل التلميذ مع عمره العقلي، ويحسب على النحو التالي:

$$\text{نسبة التحصيل (أو حاصل التحصيل)} = \frac{\text{العمر التحصيلي}}{\text{العمر العقلي}} \times 100$$

تمدّنا معايير العمر بطريقة عملية وسهلة في تفسير أداء الفرد من خلال مقابلة هذا الأداء بأداء المجموعة العمرية التي ينتمي إليها. ومعايير العمر هي من نوع المعايير الطولية التي لا تقف عند المجموعة العمرية الواحدة التي ينتمي إليها الفرد بل تتطلب العودة إلى المجموعات العمرية السابقة واللاحقة له، أي إنّها تستخرج من مجموعات عمرية متلاحقة. وكما سيتّضح لاحقاً في معرض الحديث عن التقنين ووضع الدرجات على مقياس ستانفورد – بينيه فإن حساب العمر العقلي للمفحوص بالطريقة المتبعة في هذا المقياس يرتكز على تتبّع أداء مجموعات عمرية متتابعة ويستلزم حساب العمر القاعدي أو الأساسي لهذا المفحوص (وهو العمر الذي يجيب المفحوص عن الأسئلة المخصصة له بكاملها)، كما يستلزم حساب "أوزان" أسئلة المستويات العمرية اللاحقة التي تعطى للمفحوص بصورة تدريجية وصولاً إلى العمر الأقصى أو السقف العمري لهذا المفحوص (وهو العمر الذي يعجز عن الإجابة عن أي من أسئلته).

تصلح المعايير العمرية للسمات التي تنمو مع العمر ولكنها لا تصلح لغيرها من السمات التي لا تتأثر بالعمر بصورة واضحة كسمات الشخصية مثلاً.

ويعـاب علـى المعـايير العمريـة أنهـا تعجـز عـن إمـدادنا بوحـدات قيـاس متسـاوية ومنتظمة، فالزيادة في الطول بين السنة الثالثة والسنة الرابعة من العمر مثلاً لا تعادل الزيادة في الطول بين السنة التاسعة والسنة العاشرة. كما أن سرعة النمو العقلي في الطفولة تختلف بشكل واضح عنها في فترة المراهقة. فالطفل الذي يتخلّف بمقدار سنة واحدة عن أبناء عمره وهو في الرابعة سيتخلّف بمقدار (3) سنوات وهو في عمر (12) سنة، اي إنّ النمو العقلي للطفل خلال سنة واحدة (من عمر (3) إلى (4) سنوات) يعادل نموه خلال (3)سنوات (وهو من عمر (9) إلى (12) سنة). ولعل هذا الأمر بالذات هو مما قلّل من اهمية المعايير العمرية ودعا إلى استخدام أنواع أخرى من المعايير يمكن من خلالها تخطي العيب الأكبر للمعايير العمرية والمتمثل في عدم تساوي وتائر أو معدلات النمو في الأعمار المتعاقبة.

2) معايير الصف:

لا تختلف معايير الصف أو معايير الفرقة الدراسية، كما تسمى عادة، من حيث جوهرها عن المعايير العمرية، فهي أيضاً معايير طولية ترتكز إلى صفوف (أوفرق) دراسية متلاحقة، كما تعتمد مستوى المتوسط أساساً في الحكم على أداء الفرد وتتيح مقارنته بهذا المستوى.فإن كان المتوسط الحسابي لدرجات عينة من طلاب الصف الثاني الثانوي في اختبار الكيمياء هو (30) درجة مثلاً أمكن القول ببساطة: إن الطالب الذي يحصل على الدرجة (30) يكون في مستوى المتوسط، في حين أن الطالب الذي يحصل على درجة تزيد على (30) يكون اعلى من المتوسط. أما الطالب الذي يحصل على درجة أقل من (30) فإنه سيكون أدنى من المتوسط.

ومع أن معايير الصف توفّر طريقة عملية وسهلة لتفسير أداء الأفراد تماماً كمعايير العمر، فإن معايير الصف تعاني بدورها من العيوب ذاتها التي تعاني منها معايير العمر التي تتمثل في عجزها عن إمدادنا بوحدات قياس متساوية بسبب من اختلاف وتيرة أو معدل النمو والتقدّم في الدراسة من صف

لآخر ومن مرحلة دراسية لأخرى. فليس هناك ما يسوّغ القول: إن نمو الطلاب في التحصيل في كل من المجالات الدراسية المختلفة يسير بصورة منتظمة ومطردة، أو إن الطلاب يقطعون "مسافة" واحدة في إنجازهم الدراسي من صف لآخر أو من مرحلة دراسية لأخرى. ومن الواضح أن مقدار ما يتعلّمه المرء في مجال دراسي ما لمدة سنة في صف معين قد لا يعادل ما يتعلّمه في المجال لنفسه لمدة سنة في صف آخر أعلى أو أدنى منه. (للتوسّع في المعايير الصفية والعمرية انظر:ثورانديك وهيجن، الترجمة العربية، 1989).

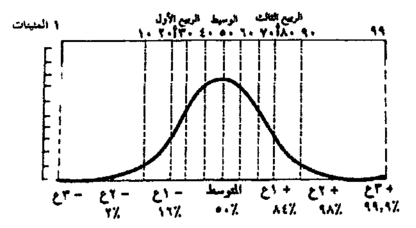
3) المعايير المئينية:

لاشك أن ترتيب درجات الأفراد باتجاه تصاعدي أو تنازلي يتيح معرفة المركز أو الموقع الذي يحتله كل منهم بين أقرانه. غير أن استخدام هذا الترتيب في تقدير أداء الفرد قد يعطي صورة مضللة عن هذا الفرد نظراً لأنه يتأثر بقوة بعدد أفراد المجموعة الذين أخضعوا لهذا الترتيب. وعلى سبيل المثال فإن من يحتل المركز الرابع في مجموعة مؤلفة من اربعة افراد يختلف بالتأكيد عمن يحتل المركز الرابع في مجموعة مؤلفة من (100) فرد. لهذا السبب عمد الباحثون إلى استخدام المئين الذي يحدّد مباشرة مركز الفرد في المجموعة معتمداً في ذلك على تحديد النسبة المئوية للحالات (او الدرجات) التي تقع تحته في التوزيع. وقد أطلق على المعيار المئيني هذا الاسم نظراً لأنه يحول التوزيع التكراري إلى أجزاء من مئة، أو أجزاء مئوية، ويقسم مستويات الأفراد إلى مئة مستوى. وعلى هذا فإن المئين (30) مثلاً هو النقطة التي يقع تحتها 30٪ من الحالات (أو الدرجات)، والمئين (10) هو النقطة التي يقع تحتها 10٪ من الدرجات (وهو العشير)، والمئين (25) هو النقطة التي يقع تحتها 25٪ من الدرجات (وهو الربيع الأول)، والمئين (50) هو النقطة التي يقع تحتها 50٪ من الدرجات (وهو الربيع الثاني أو الوسيط) وهكذا.

والمعايير المئينية هي من نوع المعايير المستعرضة أو الأفقية حيث تقوم على دراسة تشتّت درجات الأفراد الذين ينتمون إلى الصف الواحد أو العمر الواحد في الاختبار بعد ترتيبها تصاعدياً أو تنازلياً ثم بيان موقع الفرد في هذا التشتّت ومدى انحرافه عن المتوسط صعوداً أو هبوطاً. وقد يستخدم مصطلح المئينات ومصطلح الدرجات المئينية بمعنى المعايير المئينية.

تتميّز المعايير المئينية ببساطتها وسهولة حسابها ووضوح دلالتها حيث يمكن أن يفهمها الشخص العادي بسهولة إذا قيل له مثلاً: إن زيداً تفوق على 70٪ من زملائه في الرياضيات وعلى 50٪ منهم في اللغة و40٪ منهم في الجغرافية. وتصلح المعايير المئينية للأطفال والراشدين كما تصلح للاستخدام في مقاييس الشخصية بأنواعها إضافة إلى مقاييس القدرات والتحصيل، وتتفوّق بذلك على المعايير العمرية والصفية التي تصلح لمقاييس القدرات والتحصيل على وجه الخصوص.

غير أن المعايير المئينية تعاني بدورها من العيب ذاته الذي تعاني منه المعايير العمرية والصفية والمتمثل في عدم تساوي وحداتها على منحني التوزيع. فالمسافة الفاصلة بين المئين (40) والمئين (50) مثلاً تعادل تقريباً سبع المسافة الفاصلة بين نقطة الصفر والمئين (10) في المنحني الاعتدالي. ويعود السبب في ذلك إلى أن أكثر الحالات (أو الدرجات) تتجمع قرب نقطة الوسط (أو الوسيط) على هذا المنحني وأن الأقلية فقط من الحالات تنتشر بالتساوي على الطرفين (انظر الشكل رقم(3)) (نقلاً عن: Anastasi,1982 ,p.76).



الشكل رقم (3): الوحدات المئينية في التوزيع الاعتدالي

وهكذا فإنه نظراً لأن الوحدات المئينية أو المئينات تصغر في منطقة الوسط وتكبر عند الأطراف، فإن الفروق الواقعة في منطقة الوسط يلزمها – إذا صح التعبير – عدد أكبر من المئينات للتعبير عنها وتمثيلها بحجمها الحقيقي في حين أن الفروق الواقعة في الأطراف يلزمها – إذا صح التعبير هنا أيضاً – عدد أقل من المئينات لتمثيلها والتعبير عنها بحجمها الحقيقي. ولهذا فإن المئينات تبالغ في تباين الأفراد و"التفريق" بينهم كلما اقتربنا من مركز الوسط في التوزيع وتتضاءل حساسيتها للفروق الفردية على طرفي التوزيع، أو إنها بعبارة أخرى، "تغالي" في الفروق القريبة من نقطة الوسط وتقلّص أو تقلّل الفروق البعيدة عن نقطة الوسط.

بالإضافة إلى ما سبق، مع أن المعايير المئينية تكشف عن موقع الفرد بين أقرانه في المجموعة وتعطي صورة صادقة عن مركز هذا الفرد أو رتبته، وهو أمر بالغ الأهمية بطبيعة الحال، فإن هذه المعايير لا تحدّد مقدار الفرق بين درجة فرد ودرجة فرد آخر في الخاصية المقيسة. وهذا يعني أن المعايير المئينية تعجز عن إظهار الفروق بين الرتب مع أن هذه الفروق ليست منتظمة بالضرورة، أو إنها بعبارة أخرى، لا تحدّد المسافات الحقيقية الفاصلة بين الرتب المختلفة. ولا يكتفى عادة عند استخدام الرتب المئينية بعينة تقنين واحدة بل لابد من استخدام عينات تقنين متعدّدة للأعمار أو الصفوف الدراسية المختلفة (الغريب، 1970).

4) الدرجات المعيارية:

تتشابه الدرجات المعيارية مع المئينات في أنها تعتمد على دراسة تشتّت الدرجات في إطار المجموعة العمرية أو الصفية الواحدة وتنتمي بذلك إلى فئة المعايير المستعرضة أو الأفقية غير أنها تحاول أن تستدرك العيب الأكبر الذي تعاني منه المعايير المئينية وهو عدم تساوي الوحدات المئينية على منحني التوزيع. وتعدّ الدرجات المعيارية تبعاً لذلك الأداة الأفضل والأدق لتحويل (أو ترجمة) الدرجات الخام إلى مقابلاتها المعيارية. ويتيح تحويل (أو ترجمة) الدرجات الخام إلى درجات معيارية الكشف عن مدى ابتعاد (أو انحراف) درجة الفرد عن المتوسط صعوداً أو هبوطاً وتحديد مركزه مباشرة على منحني التوزيع في ضوء تمركز وتشتّت أفراد المجموعة التي ينتمي إليها.

ويمكن تعريف الدرجة المعيارية على أنها عدد وحدات الانحراف المعياري (أو الانحرافات المعيارية) التي تبعدها درجة ما عن المتوسط، وتتطلب عملية حسابها نسبة انحراف درجة الفرد عن المتوسط إلى الانحراف المعياري للدرجات أي انحراف درجات أفراد المجموعة كلها عن المتوسط. وتستخدم في حسابها المعادلة التالية:

$$ذ = \frac{خ - م}{ع}$$

حيث يشير الحرف (ذ) إلى الدرجة المعيارية (الذالية).

(خ) إلى الدرجة الخام للفرد.

(م) إلى المتوسط.

(ع) إلى الانحراف المعياري.

فإذا كانت الدرجات الخام التي حصل عليها ثلاثة من المفحوصين في اختبار المفردات هي (33) لأحدهم و(25) للثاني و(29) للثالث وكان متوسط الدرجات في المجموعة التي طبق عليها الاختبار هو (29) والانحراف المعياري لها هو (4) فإن الدرجة المعيارية التي تقابل الدرجة الخام للمفحوص الأول هي:

$$1 = \frac{29-33}{4}$$

والدرجة المعيارية التي تقابل الدرجة الخام للمفحوص الثاني هي:

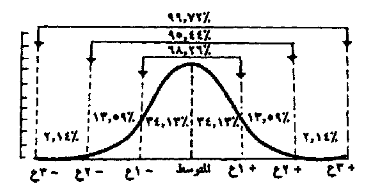
$$1- = \frac{29-25}{4}$$

في حين أن الدرجة المعيارية التي تقابل الدرجة الخام للمفحوص الثالث هي:

$$\text{صفراً} = \frac{29-29}{4}$$

وهذا يعني أن الدرجة التي حصل عليها المفحوص الأول تبعد بمقدار +1 درجة معيارية (1سيغما) عن المتوسط، وأن الدرجة المعيارية التي حصل عليها المفحوص الثاني تبعد بمقدار – 1 درجة معيارية (– 1 سيغما) عن المتوسط في حين أن الدرجة المعيارية التي حصل عليها المفحوص الثالث وقدرها صفر تقع في نقطة الوسط تماماً. ومن الواضح أن الدرجات المعيارية المعادلة للدرجات الخام التي تقع تحت المتوسط تحمل إشارات سالبة، وأن الدرجات المعيارية المعادلة للدرجات الخام التي تقع فوق المتوسط تكون موجبة، في حين أن الدرجة المعيارية المعادلة للدرجات الخام التي تقع في نقطة الوسط تعادل الصفر. وتتراوح الدرجات المعيارية التي تقع على منحني التوزع الطبيعي أو الاعتدالي، والذي تعتمده الاختبارات النفسية والتربوية، كما هو معلوم، من +3 إلى – 3 مروراً بالصفر والذي يشير إلى نقطة الوسط تماماً في التوزيع. ويظهر في الشكل التالي توزع

الدرجات المعيارية (الذالية) والنسبة المئوية للحالات التي تأخذها كل منها على منحني التوزيع الطبيعي أو الاعتدالي (الشكل رقم (4)):



الشكل رقم (4): توزع الدرجات المعيارية (الذالية) على المنحني الاعتدالي

ويتّضح من الشكل السابق أن المساحات الثلاث على الجانب الأيمن من المنحني (أي من صفر إلى +3) تغطي 49.86٪ من مجموع الحالات وأنها تعادل تماماً المساحات الثلاث الواقعة على الجانب الأيمن (أي من صفر إلى – 3). والمساحات الست مجتمعة تغطي بالتالي 99.72٪ من الحالات (أو الدرجات) مع الإشارة إلى أن نسبة ضئيلة جداً من الحالات وقدرها 0.26٪ وهي النسبة التي تهمل عادة في القياسات النفسية والتربوية، تبقى خارج الحدود المذكورة أي خارج المسافة الفاصلة بين – 3 و+3 درجات معيارية وهي تتوزع بالتساوي على جانبي المنحني (أي بنسبة 0.13٪ على كل جانب). كما يتّضح من الشكل السابق أن الأفراد يتوزعون بصورة اعتدالية أو طبيعية حيث يقع أكثر من ثلثيهم بقليل في منطقة الوسط (من صفر إلى – 1 يقع 34.13٪ من الأفراد ومن صفر إلى +1 يقع أيضاً 34.13٪ من الأفراد ويؤلفون بالتالي 68.26٪ من مجموع الحالات)، في حين أن المتطرفين قليلاً (وهم مجموع الأفراد الواقعين بين – 1 و – 2 درجة معيارية وبين +1 و+2 درجة معيارية) يؤلفون 27.18٪ من مجموع الحالات، وأن المتطرفين بشدّة (وهم مجموع الأفراد الواقعين بين – 2 و – 3 وبين +2و+3 درجة معيارية) يشكّلون نسبة ضئيلة جداً وقدرها 2.14٪ على الجانب الأيمن و2.14٪ على الجانب الأيسر اي 4.18٪ فقط من مجموع الحالات.

ومن الواضح أن الدرجة المعيارية تعتمد على المتوسط والانحراف المعياري معاً بخلاف المعايير السابقة التي تعتمد على المتوسط فقط. ولعل الميزة الأكبر للدرجات المعيارية هي أنها تمدّنا بوحدات قياس متساوية ومنتظمة وتحمل معنى واحداً من اختبار لآخر. وهي بذلك تمثّل أساساً معيارياً للمقارنة يتيح تفسير الدرجات الخام من خلال نسبة انحراف كل منها عن المتوسط إلى الانحراف المعياري للمجموعة.

غير أن الدرجات المعيارية لا تصلح للاستخدام إلا إذا كان التوزيع اعتدالياً أو أقرب ما يكون إليه، كما تتطلب استخدام إشارتي الزائد والناقص نظراً لأن نصف الدرجات المعيارية في التوزيع الاعتدالي هي درجات سالبة تقع تحت متوسط الدرجات في العينة (وهو درجة الصفر)، ونصفها الآخر هي درجات موجبة تقع فوق متوسط الدرجات في العينة (أي فوق الصفر). كما أن الدرجات المعيارية كثيراً ما تتطلب استخدام الكسور العشرية أو الفواصل ناهيك عن قيمة الصفر الموجودة في وسطها مما يضع بعض الصعوبات أمام استعمالها. ولمواجهة هذه الصعوبات عمد المعنيون إلى اقتراح بعض الدرجات المعدلة للدرجة المعيارية الأساسية (أو الذالية).

الدرجات المعيارية المعدلة:

وهي توزيع افتراضي جديد للدرجة المعيارية الأساسية (أو الذالية)، ويتلخّص الهدف من ورائه في مواجهة الصعوبات التي تعاني منها هذه الدرجة واستبعاد نقائصها. ويتحقق ذلك عن طريق التخلص من كسور أو فواصل الدرجات المعيارية بضربها بعدد ثابت كالعدد (10) مثلاً، وكذلك التخلص من الإشارات السالبة وقيمة الصفر الموجودة في وسطها بإضافة عدد ثابت كالعدد (50) مثلاً مما يتيح التعبير عنها جميعاً بأعداد موجبة وصحيحة. ومن أنواع الدرجات المعيارية المعدلة الدرجات التالية:

الدرجة التائية:

تنسب الدرجة التائية T.Score أو المعيار التائي إلى إدوارد ثورندايك ولويس ترمان، وترتكز على توزيع افتراضي للدرجات متوسطه (50) وانحرافه المعياري (10)، وتحسب باستخدام المعادلة التالية:

الدرجة التائية (ت) = الدرجة المعيارية الذالية (ذ) × 10 +50

وللدرجة التائية أهمية فائقة لكونها تصلح للاستخدام في المقاييس النفسية والتربوية كافة ويشيع استخدامها بصورة خاصة في مقاييس الشخصية.

هذا وقد عدّلت الدرجة التائية بدورها سعياً وراء رفع مستوى حساسيتها للفروق الدقيقة بين الأفراد وزيادة قدرتها على المقارنة. ومن هذه التعديلات الدرجة التائية العسكرية والدرجة التائية الجامعية.

الدرجة التائية العسكرية:

وترتكز هذه الدرجة على توزيع افتراضي للدرجات متوسطه (100) وانحرافه المعياري (20)، وقد شاع استخدامها في اختبارات التصنيف العام للجيش الأمريكي Army General Classification Test (AGCT) خلال الحرب العالمية الثانية. وتحسب باستخدام المعادلة التالية:

الدرجة التائية العسكرية = الدرجة المعيارية الذالية (ذ) ×20 +100

الدرجة التائية الجامعية:

ويطلق عليها اسم الدرجة المؤهّلة للقبول في الجامعات (CEEB) College Entrance Examination Board. وهي درجة تائية معدلة تستهدف مواجهة مشكلة الكم الهائل من الكسور العشرية التي تظهر عند استخدام الدرجات التائية الأساسية في تحديد مستويات القبول في الجامعات. وتتيح بالتالي الكشف عن مستويات الفروق الطفيفة جداً بين الأفراد باستخدام أعداد صحيحة وبأقل قدر ممكن من الكسور العشرية (أو الفواصل). وتقوم الدرجة التائية الجامعية على إخضاع الدرجات المعيارية لتوزيع جديد متوسطه (500) وانحرافه المعياري (100). والمعادلة الخاصة بها هي:

الدرجة التائية الجامعية = الدرجة المعيارية الذالية (ذ) × 100 +500

الدرجة الموزونة:

الدرجة الموزونة هي درجة معيارية معدلة تستهدف توحيد أوزان درجات الاختبارات الفرعية التي قد يضمّها الاختبار الكلي الواحد أي جعلها من وزن واحد بغض النظر عن النهاية العظمى لدرجات كل من هذه الاختبارات. وتقوم الدرجة الموزونة على توزيع متوسطه (10) وانحرافه المعياري (3)، وتحسب باستخدام المعادلة التالية:

الدرجة الموزونة = الدرجة المعيارية الذالية (ذ) × 3 + 10

ويشيع استخدام الدرجات الموزونة في اختبارات الذكاء التي اعتمدت المعايير غير الطولية ومنها اختبارات فكسلر.

الدرجة الجيمية:

تنسب الدرجة الجيمية G.Score إلى جيلفورد، وهي أيضاً من الدرجات المعيارية المعدلة. وتقوم على إخضاع الدرجات المعيارية لتوزيع جديد متوسطه (5) وانحرافه المعياري (2)، ومعادلتها هي:

الدرجة الجيمية = الدرجة المعيارية (الذالية)× 2 + 5

ومن ميزات الدرجة الجيمية أنها تصلح لقياس الفروق الكبيرة بين الأفراد حيث تقلّص هذه الفروق إلى عدد ضئيل نسبياً من المستويات من خلال تقسيم التوزيع إلى (11) قسماً (أو مستوى) يبدأ تدريجه من (صفر) إلى (10).

الدرجة الإعشارية (أو العشير):

وتمثّل الدرجة الإعشارية (أو العشير) Sten Score تعديلاً بسيطاً على الدرجة الجيمية السابقة فهي لا تختلف عنها إلا في أنها تخضع الدرجات المعيارية لتوزيع متوسطه (5.5) بدلاً من (5) وانحرافه المعياري (2)، ومعادلتها هي:

الدرجة الإعشارية (العشير) = الدرجة المعيارية الذالية (ذ) × 2 +5.5

وقد استخدمت هذه الدرجة في استخبار عوامل الشخصية لكاتل.

التسيع أو المعيار التساعي:

ويمثّل التسيع Stanine أيضاً تعديلاً بسيطاً على الدرجة الجيمية السابقة. وقد أجري هذا التعديل من قبل سلاح الطيران الأمريكي في أثناء الحرب العالمية الثانية. وهو يقوم على تقسيم التوزيع إلى (9) مستويات بدلاً من (11) مستوى وذلك بضم الدرجة (ج) صفر والدرجة (ج)1 في التسيع (1) والدرجة (ج)9 والدرجة (ج) 10 في التسيع (9). وتظهر في الجدول التالي الدرجات

التساعية وما يقابلها من نسب مئوية للحالات تحت منحني التوزيع الاعتدالي (الجدول رقم(12)):

الجدول رقم (12): الدرجات التساعية وما يقابلها من نسب مئوية للحالات:

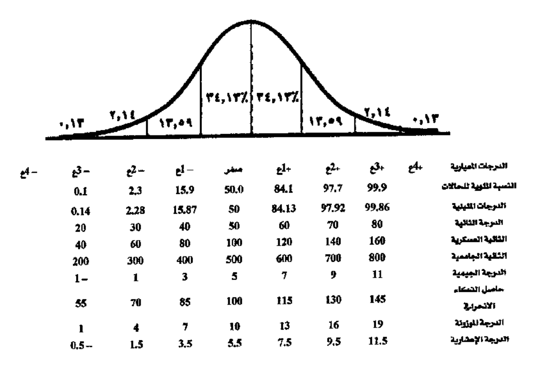
الدرجة التساعية	النسبة المئوية للحالات تحت منحني التوزع الاعتدالي
1	٪4
2	٪7
3	٪12
4	٪17
5	٪20
6	٪17
7	٪12
8	٪7
9	٪4

حاصل الذكاء الانحرافي:

ويقابل مصطلح Deviation I.Q باللغة الإنكليزية، وهو درجة معيارية معدلة تخضع الدرجات المعيارية (الذالية) لتوزيع جديد متوسطة (100) وانحرافه المعياري (15) بحسب مقياس فكسلر و(16) بحسب مقياس ستانفورد – بينيه. والمعادلة الخاصة بحساب حاصل الذكاء الانحرافي هي:

حاصل الذكاء الانحرافي = الدرجة المعيارية الذالية (ذ) × 15 (او 16) + 100

وتظهر في الشكل التالي أنواع من المعايير السابقة على المنحني الطبيعي أو الاعتدالي (الشكل رقم (5)):



الدرجات المعيارية	+4ع	+3ع	+2ع	+1ع	صفر	- 1ع	- 2ع	- 3ع	- 4ع
النسبة المئوية للحالات		99.9	97.7	84.1	50.0	15.9	2.3	0.1	
الدرجات المئينية		99.86	97.92	84.13	50	15.87	2.28	0.14	
الدرجة التائية		80	70	60	50	40	30	20	
التائية العسكرية		160	140	120	100	80	60	40	
التائية الجامعية		800	700	600	500	400	300	200	
الدرجة الجيمية		11	9	7	5	3	1	1 -	
حاصل الذكاء الانحرافي		145	130	115	100	85	70	55	
الدرجة الموزونة		19	16	13	10	7	4	1	
الدرجة الإعشارية		11.5	9.5	7.5	5.5	3.5	1.5	0.5 -	

الشكل رقم (5): انواع من المعايير على المنحني الاعتدالي

ويمكن استناداً إلى الشكل السابق وباستخدام المعادلات الإحصائية اللازمة تحويل المعايير من نوع إلى آخر وتحديد المقابلات المعيارية المختلفة لكل منها. فالدرجة المعيارية (صفر) على سبيل المثال تعادل المئين (50) والدرجة الجيمية (5) والدرجة الموزونة (10). والدرجة المعيارية (+1)، وهي تزيد بمساحة قدرها 34.13٪ على نقطة الوسط في التوزيع، تعادل المئين 84.13 (اي 50+34.13 = 84.13) كما تعادل الدرجة التائية (60) (اي 1×10+50=60) والدرجة الجيمية (7) (اي 1×2+5=7) كما تعادل الدرجة الموزونة (13) اي (1×3+10 =13) وحاصل الذكاء الانحرافي (115) اي (1×15 +100=115) وهكذا....

ويحتوي دليل الاختبار المقنن عادة قوائم أو جداول خاصة بالدرجة الخام التي يمكن الحصول عليها من الاختبار والدرجات المقابلة لها والمستخرجة من عملية التقنين (أو التعيير) كما أسلفنا. وهذه الأخيرة تمثّل قائمة (أو قوائم) المعايير التي يزوّد بها الاختبار عادة، وفي كثير من الأحيان لا تقتصر قائمة المعايير على نوع واحد من المعايير بل تتضمن أنواعاً مختلفة منها تكشف بنظرة

واحدة المقابلات المعيارية المختلفة لكل من الدرجات الخام التي يمكن أن يحصل عليها المفحوصون، وتتيح بالتالي الاطلاع على أشكال شتى من المعايير يمكن من خلالها تكوين صورة أكثر وضوحاً عن دلالة الدرجات الخام.

وتجدر الإشارة إلى أن كثيراً ما يتمّ التعبير عن المعايير بمستويات معينة يدلّ كل منها على فئة أو مجموعة معينة من الدرجات. وفي هذه الحالة لابد من تجميع الدرجات المتقاربة في فئات بهدف التعبير عنها بلغة المستويات، أو ترجمتها إلى مستويات، على أن يتمّ إخضاعها لتدريج معين كالتدريج التصاعدي الذي يبدأ بالأدنى وينتهي بالأعلى، والتدريج التنازلي الذي يبدأ بالأعلى وينتهي بالأدنى، مع تقديم أوصاف تفسيرية لها كما في الجدول التالي (الجدول رقم (13)):

الجدول رقم (13): المستويات ودلالتها:

الدرجة	المستوى	دلالة الدرجة
صفر – 14	أ	متدنية جداً
15 – 19	ب	متدنية
20 – 24	ج	متوسطة
25 – 29	د	مرتفعة
30 – 40	هـ	مرتفعة جداً

ومن الواضح استناداً إلى الجدول السابق أن فئات الدرجات قد تكون متساوية كما هي الحال في الفئات الثلاث الواقعة في وسط الجدول السابق وهي: الثانية والثالثة والرابعة، وقد لا تكون متساوية كما هي الحال في الفئتين الواقعتين على طرفي التوزيع وهما: الأولى والخامسة في هذا الجدول.

المعايير والصفحة النفسية:

بين الفوائد المهمة للمعايير أنها تمدنا بأساس موحد للمقارنة بين الخصائص أو مظاهر السلوك التي يتصدّى لها اختبار أو مجموعة اختبارات يمكن أن يخضع لها الفرد الواحد أو تخضع لها مجموعات من الأفراد. ويشير رثورندايك وإ. هيجن في هذا السياق إلى أن الأنواع المختلفة من المعايير توفّر لنا طريقة للتعبير عن الدرجات المتحصّلة على اختبارات مختلفة تماماً بوحدات مشتركة يمكن عن طريقها إجراء مقارنة ذات معنى بين هذه الاختبارات. وهذا ما يتيح رسم الصفحة النفسية (أو البروفيل) Profile للفرد الواحد أو لمجموعة من الأفراد التي تمثّل ببساطة"مجموعة العلامات المتحصّلة على اختبارات مختلفة معبراً عنها بوحدات قياس مشتركة " (ثورندايك وهيجن، الترجمة العربية، 1989، ص124). وعلى هذا فإن الصفحة النفسية (أو البروفيل) ما هي إلا شكل من أشكال التمثيل البياني للدرجات المتحصّلة على مجموعة من الاختبارات المنفصلة أو مجموعة من الاختبارات الفرعية التي يمكن أن يتألف منها الاختبار الكلي الواحد. ويتمثّل الهدف الرئيس من وراء إعداد الصفحة النفسية في التعبير عن الدرجات المتحصّلة على مجموعة من الاختبارات بطريقة واضحة للعيان وذلك عن طريق الرسم البياني الذي يأخذ شكل المنحني عادة. ويمكن عن طريق الصفحة النفسية تعرّف دلالة درجات المفحوص بالنسبة لمعيار أو أكثر من معايير الفئة أو المجموعة التي ينتمي إليها كمعايير المتوسطات والمئينات والدرجات المعيارية... إلخ. كما يمكن عن طريقها معرفة النمط أو الاتجاه العام لدرجات المفحوص وتحديد السمات أو مظاهر السلوك التي انحرف فيها عن المتوسط صعوداً أو هبوطاً. وهذا كله يتيح الكشف عن الفروق في ذات الفرد (أو المفحوص) وتعرّف جوانب الضعف والقوة في أدائه من خلال المقارنة بين الأشكال المختلفة لهذا الأداء من جهة ومقارنة كل منها بمتوسط أداء الجماعة التي ينتمي إليها الفرد من جهة ثانية. ويظهر في الشكل التالي (الشكل رقم (6)) نموذج لصفحة نفسية لدرجات أحد الأفراد على اختبارات بطارية الاستعدادات أو القدرات

الفارقية. وتظهر هذه الصفحة تفوق الفرد بوضوح في اختبار القدرة العددية (المئين 95) واختبار الاستدلال اللفظي (المئين 90) وفي اختبار القدرة العددية واللفظية معاً (المئين 97)، كما تظهر تفوقه في اختبار القواعد (المئين 85) ثم اختبار التفكير المجرد (المئين 80) ثم اختبار الاستدلال الميكانيكي (المئين 75) ثم اختبار العلاقات المكانية (المئين 70)، هذا بالإضافة إلى أنها تظهر ارتفاعه عن مستوى المتوسط بقليل في اختبارات الإملاء (المئين 55) وانخفاضه عن مستوى المتوسط في اختبار السرعة والدقة في الكتابة:

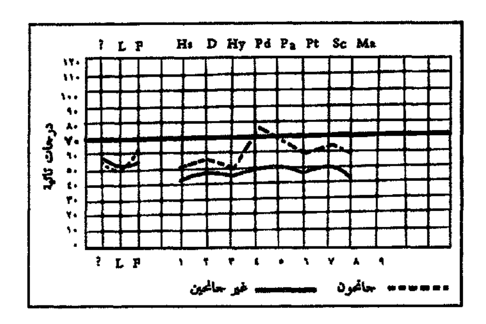
"Clerical Speed and Accuracy (المئين 40)"

الشكل رقم( ):الصفحة النفسية لأحد الأفراد على اختبارات القدرات الفارقية (DAT)
(نقلاً عن : ثورندايك و هيغن , الترجمة العربية, ١٩٨٩ ,ص١٤٦)

وتجدر الإشارة إلى أن المئينات في هذه الصفحة وزعت بحيث تكون أكثر تباعداً في الطرفين منها في منطقة الوسط وذلك للتعويض عن عدم تساوي الوحدات المئينية على منحنى التوزيع.

وثمة نماذج أخرى عديدة للصفحات النفسية منها النموذج الذي يظهر في الشكل التالي (الشكل رقم (7)) والذي يعبر عن متوسط درجات مجموعة من الجانحات ومتوسط درجات مجموعة أخرى من غير الجانحات في كل من المقاييس المختلفة التي يضمّها اختبار مينيسوتا المتعدّد الأوجه للشخصية. ويفيد

هذا النموذج في المقارنة بين الفئات أو المجموعات المختلفة (بخلاف النموذج السابق الذي يهتم بالفروق في ذات الفرد).



الشكل رقم (7): متوسط بروفيلات مجموعة من الجانحات (ن=99)
ومجموعة من غير الجانحات (ن=85)

(نقلاً عن: أنستازي، الترجمة العربية، الجزء الثاني، 1991، ص15)

ويظهر الشكل السابق بوضوح ارتفاع متوسطات درجات الجانحات في الكثير من المقاييس التشخيصية (المقاييس ذوات الأرقام: 4 و5 و6 و7 و8). وقد وصل هذا الارتفاع إلى أقصاه وتجاوز الخط الأحمر، وهو الخط الذي يبعد عن المتوسط بمقدار (2) سيغما ويشير إلى الدرجة المعيارية (70)، في مقياس الانحراف السيكوباتي (المقياس رقم (4)). كما يظهر الشكل السابق اقتراب متوسطات درجات الجانحات في المقاييس التشخيصية الثلاثة الأولى، وهي مقياس توهم المرض ومقياس الاكتئاب ومقياس الهستيريا، من متوسطات غير الجانحات.

وفي ضوء ما سبق كله يمكن أن ننتهي إلى القول: إن الصفحة النفسية تمثّل أداة ممتازة لتمثيل الدرجات المتحصّلة على الاختبارات بيانياً والتعبير عنها بدرجة قصوى من الوضوح. وتظهر فائدة الصفحة النفسية في أنها تسهّل قراءة نتائج الاختبارات وتفسيرها عن طريق إمدادنا بأساس مشترك أو موحد للمقارنة بين الدرجات من خلال المقابلات المعيارية لتلك الدرجات. وتتيح الصفحة النفسية بذلك تسليط أضواء ساطعة وقوية على تلك الزوايا والنقاط و"المفاصل" الحساسة في البنية العقلية والوجدانية للفرد، كما تتيح تشخيص جوانب الضعف والقوة في أدائه، وتعرّف الجوانب المختلفة في شخصيته، مما يسهل عمل المرشد النفسي، أو المعالج النفسي، إذا اقتضى الأمر، كما يوفّر الأساس لعملية التوجيه التعليمي والمهني وممارستها بدرجة قصوى من الفاعلية.

وثمة نقطتان لابد من أخذهما بالحسبان عند استخدام الصفحة النفسية وهما:

أ. مع أن الصفحة النفسية للفرد تمثّل أداة منظمة و"جذابة" للتعبير عن أدائه في مجالات مختلفة بدرجة عالية من الوضوح، ويمكن بالتالي أن تكشف بنظرة واحدة جوانب الضعف والقوة في هذا الأداء، فإن تفسيرها بشكل سليم يشترط أن تكون الدرجات المتحصّلة على الاختبارات المختلفة الممثّلة فيها قابلة للمقارنة. ولكي تكون هذه الدرجات قابلة للمقارنة لابد أن تشتق من أداء عينة واحدة من الأفراد كما هو الحال في الاختبارات الفرعية التي تؤلف البطارية الكلية الواحدة أو "الطاقم" التي تستخرج معاييرها عادة من أداء عينة تقنين واحدة تطبق عليها الاختبارات الفرعية جميعها التي تضمّها البطارية الكلية الواحدة. أما إذا قنّنت الاختبارات التي ستظهر نتائجها في صفحة نفسية واحدة على عينات مختلفة من حيث عددها وقدرتها التمثيلية، فإن هذا الاختلاف سيلغي الأساس الموحد الذي يمكن اعتماده في تفسير نتائج تلك الاختبارات أو يضعف فاعليته في أحسن الأحوال. وهذا ما يؤدي بالمحصلة النهائية إلى الوقوع فيما يمكن أن نسميه "ازدواجية المعايير" أو استخدام أكثر من "مكيال" واحد في تقدير أداء

الأفراد. كما يستتبع ضرورة التنبّه والحذر الشديد عند قراءة الصفحة النفسية التي تمثّل نتائج عدد من الاختبارات المنفصلة، أي التي اختص كل منها بعينة تقنين واحدة.

وسيكون من المناسب عند رسم الصفحة النفسية، تبعاً لذلك، التأكد من أن الاختبارات التي ستمثّل فيها قنِّنت على عينات متماثلة أو متكافئة إلى حد بعيد. أما عندما يكون من العسير التأكد من وجود عينات متماثلة فلن يكون هناك من مخرج سوى إعادة تقنين تلك الاختبارات على عينة مشتركة واحدة.

2. إن الفروق كلّها التي تظهر على الصفحة النفسية بما فيها الفروق الطفيفة والطفيفة جداً قد تثير الاهتمام وتكون ملفتة للنظر دون أن تحمل أية دلالة في حالات معينة. ولكي تكون للفروق دلالتها لابد أن تتمّتع بقدر من الثبات، بمعنى أنها لابد أن تظهر في حال إعادة إجراء الاختبار، كما لابد أن يكون لها أثرها في الفرد. ويشير ثورندايك وهيجن في هذا السياق إلى أن الفروق "يجب أن تكون كبيرة إلى الدرجة التي:

- نستطيع أن نكون فيها نوعاً ما متأكدين انها ستظل موجودة عند تكرار اختبار الشخص نفسه.
- يكون لها تأثير واقعي عملي.

وذلك قبل أن نبدأ بتفسير هذه الفروق واتخاذ إجراء على أساسها" (ثورندايك وهيجن، 1989، الترجمة العربية، ص147).

تحفظات يجب مراعاتها عند استخدام المعايير:

ما من شك في أن معايير الاختبارات النفسية والتربوية تنطوي على قدر كبير من الأهمية في القياسات المعيارية المرجع لكونها الأساس المعتمد في تفسير النتائج المتحصّلة وإعطائها دلالتها. ودون اعتماد معايير معينة يفترض أساساً أن

تتسم بدرجة عالية من الدقة يتعذّر إجراء مقارنات دقيقة بين جماعة وأخرى، أو بين فرد وآخر، أو بين الفرد ونفسه في تلك السمات أو تلك الجوانب من السمات (أو القدرات) التي تتصدّى لها الاختبارات النفسية والتربوية. ويمكن القول: إن العامل في القياسات النفسية والتربوية سيفتقد إلى الركيزة الأساسية من ركائز عمله، بل والمسوّغ الأساسي من مسوّغات وجوده كمختص في تلك القياسات، وقد يخبط خبط عشواء إذا افتقرت أدوات القياس التي يستخدمها إلى المعايير الإحصائية الدقيقة. ولن يكون من المناسب بالتالي اعتماد النتائج المتحصّلة من تلك الأدوات أساساً كافياً لإصدار أحكام سليمة على الأفراد أو اتخاذ قرارات حاسمة بشأنهم يمكن أن تنعكس مباشرة على مستقبلهم الدراسي والمهني، وتكيفهم الشخصي والاجتماعي، وقد تحدّد مصيرهم الحياتي برمته.

غير أن الأمر الذي يتعين على العامل في القياسات النفسية والتربوية، أو غيره ممن تهمّه تلك القياسات، أخذه بالحسبان وعدم إغفاله في أي حال من الأحوال هو أن المعايير ليست من نوع الأشياء "الثابتة" التي لا تتغير، أو إنّها منزلة أو منزّهة عن الخطأ. والأصح أن نقول بدلاً من ذلك: إنها تنطوي على الكثير من احتمالات الخطأ، وإن لها محاذيرها وعيوبها. ويؤدي استخدامها إلى إعطاء صورة مضللة عن الأفراد إذا افتقرت إلى الدقة اللازمة، بل قد تكون على درجة عالية من الدقة بحد ذاتها، ويُساء مع ذلك استعمالها، وتعمل بالاتجاه المعاكس للاتجاه الذي أعدّت له أصلاً. ويمكن إجمال النقاط التي لابد من الوقوف عندها، والتحفظات التي لابد من أخذها بالحسبان عند استخدام المعايير الإحصائية النفسية والتربوية فيما يلي:

1. إن معايير الاختبارات النفسية والتربوية لا تهبط علينا من السماء بل تستخرج من أداء مجموعة محددة من الأفراد تمّ اختيارها لتكون عينة ممثلة للمجتمع. ومن هذه الزاوية تخضع معايير الاختبارات النفسية والتربوية مباشرة لعينة التقنين، وللشروط الواقعية المحيطة بأداء أفرادها، ويتحدّد صدقها بصدق هذه العينة وقدرتها على تمثيل المجتمع الأصلي

الذي سُحبت منه. وتؤكد تبعيّة المعايير للعينة ولشروط الزمان والمكان التي يعمل فيها أفرادها الطبيعة النسبية للمعايير وعدم جواز النظر إليها وكأنها من نوع الأشياء الثابتة التي لا تتغير أو التي تنعدم فيها احتمالات الخطأ كما أسلفنا.

2. يمثّل المعيار الأساسي السائد في القياسات النفسية والتربوية مستوى المتوسط في المجموعة العمرية أو الصفية ويعتمد أساساً للمقارنة والحكم. ومن الواضح أن المتوسط بحد ذاته ليس أنموذجاً يُحتذى به، بل هو قيمة كمية إحصائية تعبر عن أداء المجموعة ككل وبما تضمّه من أفراد يقترب حوالي ثلثيهم في التوزيع الاعتدالي أو السوي الذي تنطلق منه المعايير (68.26٪) من مركز المتوسط، ويتوزع الثلث الباقي (31.74٪) على طرفي التوزيع مناصفة ليؤلف الفئتين العليا والدنيا. وهذا يعني بعبارة أخرى أنّ المتوسط ما هو إلا صورة رقمية تمثّل المجموعة بكل أفرادها بدءاً بأدناهم وانتهاء بأعلاهم مروراً بالمتوسط بينهم. ومن هذه الناحية تُعبّر المعايير في القياسات النفسية والتربوية عما هو كائن ولا تُعبر عما يجب أن يكون كما هو الحال بالنسبة للمعايير الخلقية والاجتماعية، إنها تمثّل وصفاً للوضع القائم كما يؤكد ثورندايك وهيجن، وليست مواصفات لما يجب أن يكون عليه الحال. (ثورندايك وهيجن، 1989، الترجمة العربية، ص148). وهذا الأمر يؤكد بدوره الطبيعة النسبية للمعايير و"مشروطيتها" بمستويات الأداء جميعها الدنيا منها والعليا، بالإضافة إلى الوسطى، ولا يتيح بأي حال من الأحوال استخدامها بوصفها القدوة التي يُقتدى بها أو المثال الذي نسعى إليه. ومن الواضح أن مستوى المتوسط لدى مجموعة من الأفراد قد لا يتجاوز المستوى الأدنى في الخاصية المقيسة لدى مجموعة ثانية من الأفراد، وقد يتجاوز المستوى الأعلى من مستويات الأداء لدى مجموعة ثالثة.

3. يفقد معيار المتوسط أهميته بل وتنعدم قيمته تماماً عندما يتركز الاهتمام على مقارنة إنجاز الفرد بمحكات مطلقة للأداء، أو يكون الهدف

من عملية القياس مثلاً هو الكشف عن المستوى الأعلى من مستويات التعلّم، والذي يعبر عنه بمستوى الإتقان. ومن المعلوم انه لا مكان لمعايير المتوسطات في القياسات المحكّية المرجع، أو القياسات بدلالة المحكّ، التي تسعى إلى بيان الإتقان أو عدم الإتقان ولا شيء آخر سواه. غير أن هذا لا يعني أن القياس التقليدي الذي يرتكز على معايير المتوسطات، أو ما يوصف عادة بالقياس المعياري، أصبح من الأشياء التي عفا عليها الزمان. فهذا النوع من القياس ما زال يحتل مساحة واسعة على أرضية القياس النفسي والتربوي المعاصر. وستحتفظ المعايير التي يرتكز إليها بقيمتها حيثما كان الهدف هو مقارنة أداء الفرد بأداء المجموعة العمرية أو الصفية التي ينتمي إليها، أو مقارنة أداء مجموعة من الأفراد بأداء مجموعة أخرى، أو مقارنة أداء الفرد في مجال أو أكثر بأداء هذا الفرد نفسه في مجالات أخرى متنوعة بدلالة الأداء المتوسط لمجموعته.

4. المعايير التي تستخرج في فترة زمنية معينة تصلح لهذه الفترة بالذات ولا تصلح، أو لا تصلح بالدرجة نفسها، للفترة أو الفترات الزمنية اللاحقة. ومع أن الفترة الزمنية التي تصلح لها المعايير غير محددة تماماً فقد جرت العادة على ألا تتجاوز هذه الفترة خمسة عشر عاماً، وأحياناً عشرين، أو ثلاثين عاماً على أبعد تقدير في الاختبارات الواسعة الانتشار. والمعايير المستخرجة في ستينيات هذا القرن أو سبعينياته قلما تصلح، وربما لا تصلح لأيامنا هذه. وهذا يعني بعبارة أخرى: إنَّ المعايير المستخرجة تهرم وتضعف مصداقيتها بمرور الوقت، ولابد من تحديثها بعد مرور فترة زمنية معينة، وإعادة تقنين الاختبار على عينة جديدة من الأفراد تعكس صورة المجتمع بوضعه الجديد، كما تعكس التغيرات أو التحسينات المختلفة التي يمكن أن تطرأ على أداء أفراده. ومن المؤكد أن معارف أطفال العقد الأخير من القرن العشرين أفضل من معارف العقدين السابع والثامن من هذا القرن مثلاً. ويتوقع أن تتحسن معارف الأطفال في العقود القليلة القادمة بصورة واضحة بفعل الثورة المعرفية والمعلوماتية التي يشهدها عصرنا. وقد أشارت

انستازي في هذا السياق إلى أن معايير الذكاء التي يوفّرها تقنين عام 1972 لمقياس ستانفورد – بينيه اظهرت تحسناً واضحاً في الأداء الاختباري للمفحوصين في المستويات العمرية جميعها بالمقارنة مع التقنين السابق، ويبلغ هذا التحسن وسطياً (10) درجات في حاصل الذكاء لدى أطفال مرحلة ما قبل المدرسة، وهو يعود برأي أنستازي إلى تأثير طائفة واسعة من العوامل والمتغيرات بينها "انتشار وسائل الاتصال المختلفة، وارتفاع مستوى ثقافة الوالدين، وعوامل أخرى " (Anastasi,1982,p.230). وهذا كله يشير بدوره إلى الطبيعة النسبية للمعايير، وتبعيتها لشروط الزمان، ويؤكد ضرورة استخراج معايير جديدة للاختبارات من فترة لأخرى.

5. المعايير التي تستخرج من أداء عينة من الأفراد سحبت من مجتمع معين تصلح لهذا المجتمع بالذات ولا تصلح، أو لا تصلح بالدرجة نفسها، لمجتمع أو مجتمعات أخرى لابد أن تختلف في قليل أو كثير عن المجتمع الأصلي الذي أعدت له هذه المعايير. فمعايير الاختبار التي استخرجت من أداء عينة قومية أمريكية وأعدت للاستخدام في المجتمع الأمريكي لا تصلح للاستخدام في سورية مثلاً، ولابد من إعادة تقنين هذا الاختبار على عينة ممثلة لأبناء القطر العربي السوري واستخراج معايير جديدة له، وإلا فإن الاختبار لن يكون صالحاً للاستعمال في سورية، وقد يؤدي استعماله إلى إعطاء صورة مضللة عن أداء الأفراد، ويحمل من الأضرار أضعاف ما يحمله من الفوائد والخيرات.

6. لابد من التمييز بين المعايير القومية الشاملة وهي المعايير التي تُستخرج من أداء عينة قومية واسعة من الأفراد تمثّل المجتمع العريض بطبقاته وفئاته المختلفة، والمعايير المحلية أو الإقليمية، وهي المعايير الخاصة بالجماعات أو الثقافات الفرعية Sub – Cultures داخل المجتمع الواحد، التي تستخرج من أداء عينات تمثّل كل منها إحدى الثقافات أو الجماعات الفرعية داخل هذا المجتمع. والمعايير القومية هي الأكثر صلاحاً للاستخدام حين يكون الغرض من عملية القياس هو مقارنة أداء الفرد بأداء

أبنـاء عمـره، أو أبنـاء صـفه الدراسـي في البلـد عامـة، وتعـرّف مسـتوى نمـوه العقلي أو اللغوي أو تحصيله المدرسي مـثلاً بالمقارنة مـع نظرائـه في العمـر أو الصـف مـن أبنـاء البلـد بأسـرهم. والمعـايير المحليـة هـي الأكـثر صـلاحاً للاستخدام حين يكون الغرض من عملية القياس مقارنة أداء الفرد بأداء أبنـاء الإقليم أو المنطقـة التعليميـة (أو المحافظـة) الـتي ينتمـي إليهـا مـثلاً، ولاسيما حين تختلف هذه المنطقة بصورة واضحة عن بقية المناطق ضمن البلـد الواحـد مـن حيـث مسـتوى التعلـيم، ودرجـة التقـدّم الاقتصـادي والاجتماعي وغيرها. وبصورة عامة يتوقع أن تكون المعايير المحلية هي الأدق بالمقارنـة مـع المعايير القومية الشاملة نظراً للصعوبات الكبيرة التي تواجه عملية سحب عينة تقنين ممثلة لمجتمع واسع ومترامي الأطراف بطبقاته وفئاته المختلفة.

7. لابد من التمييز أيضاً بين المعايير التي ترتكز إلى أداء المجتمع بكليته، سواء أكانت مـن نـوع المعـايير القوميـة الشـاملة، أم المعـايير الإقليميـة (المحليـة)، والمعايير التي ترتكز إلى أداء واحدة أو أكثر من الفئات الخاصة التي يطلق عليهـا عـادة اسـم "معـايير الفئـات الخاصـة" فمعـايير النـوع الأول تصـلح للمجتمـع الكلـي بقطاعاته وفئاته المختلفة، ولا تصلح لتلك الفئات التي تحتـاج إلى رعايـة خاصـة كفئـات الصـم، والمكفـوفين، والمصـابين بالشـلل التوافقي، وهي الفئات التي لابد أن تُستخرج لها معاييرها الخاصة. فالحكم على أداء طفل أصم أو مكفوف في مجـال معـين أو أكثر لابد أن يقوم على مقارنة أداء هذا الطفل بمستوى المتوسط في الفئة التي ينتمـي إليهـا (فئة الصم أو المكفوفين أو غيرهم) ولـيس بمسـتوى المتوسـط لـدى مجموعـة مـن الأطفال العاديين. ومن أمثلة الاختبارات التي قننت على عينة مشتقة من فئة خاصـة مـن الأفـراد اختبـار هسـكي – نبراسـكا للاستعداد للـتعلّم الـذي استخرجت له معايير خاصة للأطفال الصم.

هذا وتجدر الإشارة إلى أن الكثير من اختبارات القدرات أو الاستعدادات الخاصة لها معاييرها الخاصة أيضاً التي تُستخرج من أداء عينات من الأفراد في المهن المختلفة التي تتطلبها تلك القدرات أو الاستعدادات. وتظهر فائدة هذا النوع الأخير من المعايير في توفير الأساس الذي يمكن اعتماده في الحكم على الفرد من خلال بيان مدى توافر تلك القدرات أو الاستعدادات الخاصة لديه واللازمة لمزاولة مهنة أو عمل معين بدرجة عالية من الكفاية. ومن الواضح أن مدارس الطيران مثلاً لابد أن تكون لها معاييرها الخاصة في انتقاء الأفراد الراغبين بالالتحاق بها.

8. ما زلنا في الوطن العربي بصورة عامة، نعاني من فقر واضح في مجال تقنين الاختبارات النفسية والتربوية. ومع أن السنوات الأخيرة من القرن الماضي والأولى من هذا القرن شهدت توسّعاً في هذا المجال وإنّ الكثير من اختبارات الذكاء والشخصية الواسعة الانتشار في العالم الغربي قُننت وأعدت للاستخدام في بعض الأقطار العربية، كمصر، والأردن، والسعودية، فإن الأمر الذي لا يختلف عليه اثنان هو أن ثمة مهمات جسيمة وأعباء كبيرة للغاية تنتظر العاملين في القياس النفسي والتربوي في الأقطار العربية سواء في مجال تقنين الاختبارات الأجنبية وتوفير معايير قومية وقطرية لها، أم في مجال إنتاج اختبارات مقننة ذات هوية عربية بحتة.

الفصل الثامن

استخدام الإحصاء في تحليل نتائج الاختبارات وتفسيرها

الفصل الثامن

استخدام الإحصاء

في تحليل نتائج الاختبارات وتفسيرها

ظهرت أهمية علم الإحصاء في أواخر القرن الثامن عشر حين تركز اهتمام العلماء من مثل لابلاس وجاوس في الدراسة الكمية للعلاقات بين الظواهر والحوادث المختلفة، وحاولوا إخضاعها للتحليل الإحصائي وقوانين الاحتمال وغيرها.

وقد بدأ الإحصاء منذ ذلك الحين يشغل حيزاً هاماً في مجالات علمية متنوعة من مثل علم الحياة والكيمياء والديمغرافيا والفلك، كما أن علوماً أخرى كثيرة (تصنف عادة في فئة العلوم الإنسانية أو الاجتماعية) من مثل علم الاجتماع وعلم الاقتصاد وعلم النفس والتربية اتجهت بدورها إلى الإفادة من المنهج الإحصائي واستخدامه في دراسة الظواهر التي تتصدى لها. ويحتل الإحصاء في الوقت الحاضر مكانة هامة في العلوم النفسية والتربوية، وتظهر أهميته بصورة خاصة في القياس النفسي والتربوي المعاصر الذي يحتاج إلى الإحصاء لمعالجة البيانات العددية المتحصلة من عملية القياس وتفسيرها. والواقع أن نتائج القياس لا معنى لها بحد ذاتها ولابد من إخضاعها للمعالجة الإحصائية ليتضح معناها وتتكشف دلالتها، ويتم بالتالي تحقيق الفائدة المرجوة منها. ولعل من المزايا الهامة لعلم الإحصاء عند تطبيقه في القياس التربوي أنه يفسح مجالاً واسعاً للمقارنة، كأن نقارن تحصيل التلميذ في مادة بتحصيله في مادة أخرى، أو نقارن تحصيل التلميذ في مادة بتحصيل زملائه في المجموعة في هذه المادة، ومن الأسئلة الهامة التي ستطرح في هذا الفصل:

1. ما معنى العلامة الخام؟
2. كيف نحلل نتائج الاختبار وتفسر؟
3. ما متوسط العلامات على الاختبار وما مدى انتشار العلامات حول هذا المتوسط؟
4. هل الذين يتفوقون في المادة (أ) يتفوقون في المادة (ب) مثلاً، وإلى أي مدى يرتبط التحصيل في المادة (أ) بالتحصيل في المادة (ب)؟
5. ما الدلالة الإحصائية للنتائج المستخرجة؟
6. ما هي أهم مقاييس الدلالة الإحصائية وما فوائدها؟
7. ما الغرض من استخدام أسلوب العينات بدلاً من المسح الشامل؟
8. كيف يمكن اختيار عينة ممثلة للمجتمع الأصلي (مجتمع الدراسة)؟
9. ما التحليل الكمي لبنود الاختبار وما فوائد هذا التحليل؟

إن هذه الأسئلة وغيرها كثير تقع في نطاق علم الإحصاء الوصفي والتحليلي. وتتضمن الأجزاء التالية من هذا الفصل محاولات للإجابة عنها بالقدر الذي تتطلبه خطة هذا الكتاب، ودون التعرض لبعض المسائل والتفصيلات الواسعة التي يمكن الرجوع إليها في كتب الإحصاء المتخصصة. والأمر الذي تجدر الإشارة إليه في هذا السياق أن العامل في القياس النفسي والتربوي، أكان باحثاً أم معلماً، لا يهمه من الإحصاء سوى أنه أداة لتفسير نتائج القياس، وأما صياغة القوانين الإحصائية ودراسة أسسها الرياضية فلا تدخل في مجال عمله بل تدخل في مجال عمل المتخصص في الإحصاء. من جهة أخرى فإن العمليات الإحصائية، على الرغم من وقعها الثقيل على بعض الأفراد، أصبحت أمراً يسيراً للغاية بعد ظهور الحاسبات الإلكترونية وانتشار الآلات الحاسبة الصغيرة (حاسبات الجيب) في كل مكان وبأسعار رخيصة. وقد تضاءلت في هذا العصر أهمية القيام بالعمليات الإحصائية بالطرائق اليدوية، ولكن مع ذلك، لابد للعامل في القياس من تعرف طبيعة العمليات الإحصائية اللازمة له حتى لو استعان بالحاسب الإلكتروني في إجراء تلك العمليات، إذ لابد أن يجد لغة

مشتركة مع مخطط البرامج، ويحدد بالاشتراك معه نوع المعالجات الملائمة للبيانات والنتائج التي يتوصل إليها.

ترتيب العلامات وجدولتها:

من الطرائق المتبعة في عرض العلامات بأسلوب مبسط وواضح ترتيبها تصاعدياً أو تنازلياً، حيث يظهر في هذا الترتيب مركز كل طالب بالنسبة لزملائه في المجموعة كما تظهر العلامة الأعلى والعلامة الأدنى في هذه المجموعة. فلو كان عدد الطلاب في الصف = 12 وكانت علاماتهم الخام في اختبار معين نهايته العظمى هي (20) كالآتي:

13، 3، 18، 5، 7، 10، 8، 10، 5، 16، 8، 10

فإن ترتيب هذه العلامات تصاعدياً أو تنازلياً يعطينا فكرة أوضح عن توزعها، كما يظهر بوضوح العلامة العليا والعلامة الدنيا بينها. ويتضح ذلك بعد أن ترتب هذه العلامات تنازلياً فيما يلي:

18، 16، 13، 10، 10، 10، 8، 8، 7، 5، 5، 3

الجدول التكراري:

إن الجدول التكراري هو صورة مبسطة ومختصرة لعرض العلامات الخام، وهو يتألف من عمودين رئيسين يبيّن الأول العلامات الخام مرتبة ترتيباً تنازلياً أو تصاعدياً، ويبيّن الثاني عدد مرات ظهور كل علامة أو تكرارها. ويبيّن الجدول التالي التوزيع التكراري لمجموعة العلامات الخام السابقة التي تم ترتيبها تنازلياً:

الجدول رقم (14): التوزيع التكراري لعلامات 12 طالباً:

التكرار	العلامة الخام (خ)
1	18
0	17
1	16
0	15
0	14
1	13
3	10
0	9
2	8
1	7
0	6
2	5
1	3
0	2
0	1

وإذا كان عدد العلامات كبيراً، فإن من الأفضل أن نعمل على تقليل عددها بأن نقوم بتجميعها في فئات، ولنأخذ المثال التالي لعلامات (52) طالباً في اختبار ما بعد أن تم ترتيبها تنازلياً:

الجدول رقم (15): علامات مجموعة من الطلاب مرتبة تنازلياً:

21	27	30	32	35	37	40	46	59
20	25	29	32	35	37	39	44	56
17	24	29	31	34	36	38	43	52
17	22	28	31	34	36	38	43	50
	22	27	31	33	36	38	42	50
	21	27	30	32	36	37	42	47

إن تحسين طريقة عرض العلامات السابقة يمكن أن يتم بتجميع بعضها إلى بعض في (15) فئة مثلاً. وفي هذه الحالة يمكن أن نجمع كل ثلاث علامات مع بعضها بحيث أنّ كل مجموعة أو فئة تمثل (3) علامات، كما لا بد أن تذكر الحدين الأدنى والأعلى للفئة ومدى الفئة. وفي المثال الحالي حيث العلامة العليا هي 59 والدنيا هي 17 فإن مدى العلامات هو: 59 – 17=42.

وإذا قسمنا 42 على 15 فإننا نحصل على 2.8، ولما كان أقرب عدد صحيح لهذه القيمة هو 3 فإننا نأخذ الرقم 3 على أنه مدى الفئة، ويصبح التوزيع المجمع في فئات في المثال الحالي على النحو التالي:

الجدول رقم (16): التوزيع التكراري لعلامات خام مجمعة في فئات لـ 52 طالباً:

الفئة	الإشارات	التكرار	الفئة	الإشارات	التكرار
60 – 58	/	1	36 – 34	///// ///	8
57 – 55	/	1	33 – 31	///// //	7
54 – 52	/	1	30 – 28	/////	5
51 – 49	//	2	27 – 25	////	4
48 – 46	//	2	24 – 22	///	3
45 – 43	///	3	17 – 19	///	3
42 – 40	///	3	18 – 16	//	2
39 – 37	///// //	7			

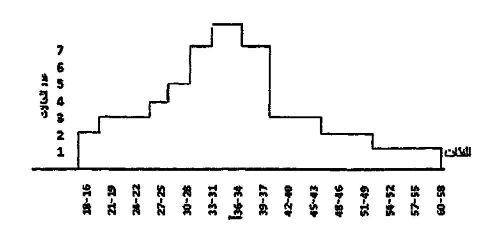
إن الجدول السابق هو جدول مختصر يظهر فيه عدد الأفراد الموجودين في كل فئة. وبإلقاء نظرة سريعة إلى هذا الجدول يتبيّن أنّ لدينا (5) أفراد في الفئة 28 – 30 و(8) أفراد في الفئة 34 – 36 وهكذا، وبالطبع ليس من الضروري في كل الحالات أن يكون مدى الفئة هو 3، وقد يصل هذا المدى إلى 5 أو 10 إذا

كان العدد كبيراً. والواقع انه ليس هناك قاعدة ثابتة بصدد عدد الفئات، ولكن أكثر ما يكون عدد الفئات في دراسة المسائل التربوية بين 6 – 15 فئة. فإذا كانت العلامات الأصيلة تغطي مدى لا يزيد على (20) علامة فإن التجميع لا ضرورة له. وعموماً فإن التوزيع التكراري، شأنه شأن جميع الطرق الإحصائية، هو وسيلة لتلخيص وتنظيم البيانات الكمية حتى تسهل معالجتها، ويعود تقدير مدى الفئة إلى الباحث نفسه أو المعلم.

التمثيل البياني:

يمكن تحويل الجدول السابق إلى رسم بياني يبيّن خواصه بصورة أوضح. ومن أنواع الرسوم البيانية المدرج التكراري الذي نحصل عليه برسم خطين أحدهما افقي والآخر عمودي بحيث يمثل الخط الأول (الأفقي) الفئات، والثاني (العمودي) التكرارات، ثم نقوم بتقسيم المحور الأفقي إلى عدد من الأقسام المتساوية يماثل عدد الفئات، وهي في مثالنا الحالي (15) فئة، كما نقوم بتقسيم المحور العمودي إلى عدد من الأقسام المتساوية يماثل اكبر تكرار في الجدول وهو (8) في مثالنا.

ويظهر ذلك في الشكل التالي رقم 8:



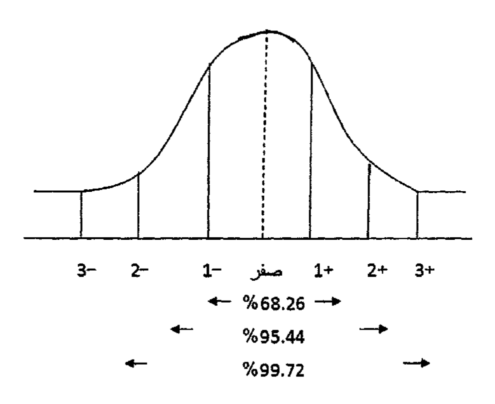
الشكل رقم (8) المدرج التكراري لدرجات 52 طالباً

ويمكن تمثيل البيانات السابقة بالمضلع التكراري الذي يقوم على الأساس السابق نفسه ويمكننا من رؤية البيانات بصورة أفضل. وفي هذه الحالة نحدد نقطة تقع فوق وسط كل فئة من الفئات بحيث يكون ارتفاعها مساوياً عدد مرات تكرار هذه الفئة، ثم نقوم بعد ذلك بوصل هذه النقاط بخطوط مستقيمة، فإذا وصلنا بين نقاط الالتقاء المتعددة حصلنا على المضلع التكراري.

وبالإضافة إلى المدرج التكراري والمضلع التكراري يمكن التعبير عن البيانات العددية بالمنحني التكراري، وهو المضلع التكراري نفسه بعد أن تتم تسويته وإزالة الزوايا منه بحيث يأخذ شكل الخط المنحني المتواصل بدلاً من الخط المنكسر، ويمر بأكثر نقاط الالتقاء ولا يكون ابتعاده عن النقاط التي لا يمر بها إلا قليلاً. والنوع الشائع للمنحني التكراري هو المنحني التكراري المعتدل أو الطبيعي أو السوي (منحنى جاوس)، وهو يدل على أن أكثرية الأفراد حصلت على قيمة حول المتوسط، وأن عدد حالات التكرار فوق المتوسط يعادل عدد حالات التكرار تحت المتوسط، كما أن عدد الحالات المتطرفة إلى اليمين أو اليسار قليلة. والمنطلق في المنحنى الطبيعي أو السوي أن الصفات البشرية كصفة الطول أو الوزن أو الذكاء أو غيرها تتوزع بصورة اعتدالية بحيث تقع الأكثرية من الناس في الوسط والأقلية تحت الوسط وفوقه. إضافة إلى ذلك فإن هذا المنحنى منتظم ومتناظر، وينطبق نصفاه انطباقاً تاماً أحدهما على الآخر، كما أن قيمة المتوسط والوسيط والمنوال واحدة فيه.

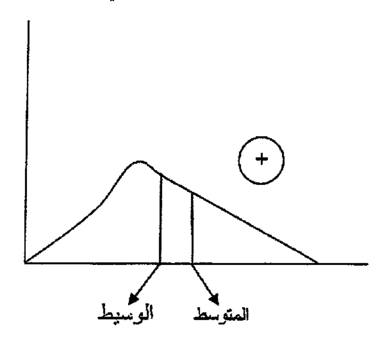
وفي الشكل التالي (رقم 2) لمنحني التوزع السوي الذي كثيراً ما يشبّه بالجرس تظهر نسبة الحالات التي تقع في منطقة المتوسط وهي تمثل أكثر من ثلثي عدد الحالات (68.26٪) وتبعد بمسافة انحراف معياري واحد فوق المتوسط أو تحته، كما تظهر نسبة الحالات المتطرفة التي تمثل أقل من ثلث عدد الحالات وتبعد بمسافة +2 و+3 و – 2 و – 3 انحراف معياري عن المتوسط.

إن الفرضية السائدة هي أن الصفات البشرية تتوزع توزعاً سوياً وتؤدي إلى المنحني السوي. والمنحنى الملتوي يوضح لنا مدى الخروج عن الاستواء فإما أن تكون المجموعة منتقاة في هذه الحالة (كالأطباء من ناحية الذكاء، أو طلاب الكلية الحربية من ناحية الطول، أو الطلاب الراسبين في صفهم من ناحية التحصيل)، او ان ثمة خطأ ما في القياس، أو أن عدد أفراد المجموعة قليل مما يدفع بها عن الاعتدال.



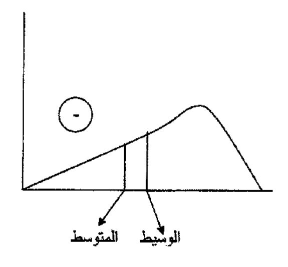
الشكل رقم (9): منحني التوزع السوي

ويشير الشكل التالي إلى توزيع ملتوٍ التواءً موجباً يلاحظ فيه نقصان التكرار بزيادة العلامة. وعادة يوصف التوزيع بأنه ملتوٍ التواءً موجباً إذا كانت درجة المتوسط أعلى من درجة الوسيط في هذا التوزيع وذلك لأن المتوسط يتأثر بالدرجات المتطرفة بخلاف الوسيط.



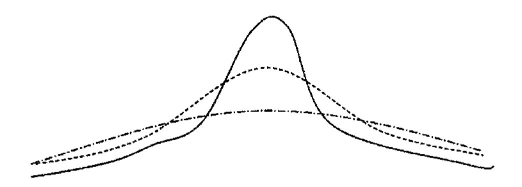
الشكل رقم (10): منحنٍ تكراري ملتوٍ التواءً موجباً

وأما الشكل التالي فيشير إلى توزيع ملتو التواءً سالباً يلاحظ فيه زيادة التكرار بزيادة العلامة (أي أن الكثير من الأفراد حصلوا على علامات عالية في هذا التوزيع). والالتواء السلبي يعني أن درجة المتوسط أقل من درجة الوسيط في التوزيع.



الشكل رقم (11): منحنٍ تكراري ملتوٍ التواءً سالباً

وبالإضافة إلى خاصية الالتواء هناك خاصية أخرى للمنحني التكراري، وهي خاصية التفلطح أو الانبساط. فالمنحنى التكراري قد لا يكون اعتدالياً وقد يظهر توزيعاً مفلطحاً أو منبسطاً بحيث يقل تمركز الدرجات في نقطة الوسط مقابل زيادة انتشارها على مدى التوزيع كله، وقد يظهر توزيعاً حاداً يتزايد فيه تمركز الدرجات في نقطة الوسط مقابل تناقص انتشارها على مدى التوزيع كما في الشكل التالي الذي يظهر توزيعاً مفلطحاً وتوزيعاً حاداً بالمقارنة مع توزيع اعتدالي أو سوي.



الشكل رقم (12): يبين توزيعاً مفلطحاً (- - -) وتوزيعاً حاداً (—) مقارناً بتوزيع اعتدالي (......)

مقاييس النزعة المركزية:

غالباً ما تفسر علامة التلميذ في اختبار معين من خلال مقارنتها مع "القيمة الوسطى للأداء في هذا الاختبار. ومن المعلوم أن أكثر العلامات في التوزيع الاعتدالي تنزع إلى التمركز في نقطة معينة هي نقطة الوسط وتمثل هذه النقطة أساساً أو معياراً للأداء تتم مقارنة علامة التلميذ به وتقدر بأنها في مستوى الوسط أو أعلى منه أو دونه. ويستخدم المربون عادة في تحليلهم وتفسيرهم لنتائج الاختبارات ثلاثة أنواع من مقاييس القيم المتوسطة أو النزعة المركزية وهي: المتوسط والوسيط والمنوال أو الشائع.

المتوسط:

إن المتوسط الحسابي المجموعة من القيم أو العلامات هو ببساطة مجموع هذه القيم أو العلامات مقسوماً على عددها. ويحسب المتوسط بالمعادلة التالية:

$$م = \frac{مج\ س}{ن}$$

حيث: يشير الحرف م إلى المتوسط

ومج س إلى مجموع القيم أو العلامات.

و ن إلى عدد القيم أو العلامات.

فمثلاً إذا أردنا حساب متوسط العلامات التالية:

9، 5، 7، 8، 2، 3، 6، 7، 5، 8

فإننا نعوض في المعادلة السابقة ونحصل على النتيجة التالية:

$$م = \frac{8+5+7+6+3+2+8+7+5+9}{10} = 6$$

ومن الواضح أنه إذا كانت القيم محدودة العدد فقد يكون هذا الأسلوب هو الأسلوب المفضل. إلا أن هذا الأسلوب يصبح مرهقاً حين يكون العدد كبيراً. وهناك أساليب أكثر فعالية بالنسبة للأعداد الكبيرة وهي مبنية على خصائص التوزيع التكراري. ومن أبسط الأساليب في حساب المتوسط لقيم منظمة في جدول تكراري الأسلوب الذي يقوم على ضرب كل قيمة بتكرارها ثم جمع حواصل الضرب وتقسيمها على عدد القيم وذلك وفق المعادلة التالية:

$$م = \frac{مج\ (س \times ك)}{ن}$$

حيث يشير الرمز:

م: إلى المتوسط.

س: إلى كل واحدة من القيم

ك: إلى التكرار

ن: إلى العدد.

وفي الجدول التالي مثال لهذه الطريقة:

الجدول رقم (17): تنظيم حساب المتوسط للقيم وتكرارها:

القيم (س)	التكرار (ك)	القيم × التكرار (س× ك)
10	3	30
9	2	18
8	4	32
7	10	70
6	8	48
5	14	70
4	12	48
3	5	15
2	2	4
1	2	2
المجموع	62	337

وبتطبيق المعادلة السابقة والتعويض عنها يكون المتوسط هو:

$$5.435 = \frac{337}{62}$$

إلا أن الطريقة السابقة غير مناسبة حين تكون الأعداد كبيرة جداً ويكون المدى كبيراً بين القيمة الدنيا والقيمة العليا. ومن الأفضل في هذه الحالة تجميع القيم في فئات ثم يحسب مركز كل فئة ويعدّ ممثلاً لها. وتطبق بهذه الطريقة المعادلة التالية:

$$\text{م} = \frac{\text{مج (س × ك)}}{\text{ن}}$$

حيث يشير الحرف:

م: إلى المتوسط

س: إلى مركز الفئة وهو نقطة الوسط بين الحدين الأدنى والأعلى للفئة.

ك: إلى التكرار

ن: إلى العدد.

وفي الجدول التالي مثال لهذه الطريقة.

الجدول رقم (18): تنظيم حساب المتوسط لفئات القيم:

فئات القيم	التكرار (ك)	مركز الفئة (س)	مركز الفئة × التكرار (س×ك)
50 – 54	2	52	104
45 – 49	3	47	141
40 – 44	5	42	210
35 – 39	10	37	370
30 – 34	6	32	192
25 – 29	5	27	135
20 – 24	2	22	44
المجموع	33		1196

وبتطبيق المعادلة السابقة والتعويض عنها يكون المتوسط هو:

$$\frac{1196}{33} = 36.24$$

وبالإضافة إلى الطرائق السابقة في حساب المتوسط ثمة طرائق أخرى تقوم على اختيار إحدى العلامات أو فئات العلامات كمتوسط فرضي. ويؤخذ المتوسط الفرضي بصورة تقديرية ويحسب المتوسط الحقيقي على أساس الفرق بين المتوسط الفرضي وبين كل قيمة من القيم، وذلك وفق المعادلة التالية:

$$\text{م} = \text{م صفر} + \frac{\text{مج}\left(\text{ح} \times \text{ك}\right)}{\text{ن}}$$

حيث يشير الحرف:

م: إلى المتوسط

م صفر: إلى المتوسط الفرضي

ح: إلى مقدار انحراف كل قيمة عن المتوسط الفرضي.

ك: إلى التكرار.

ولو طبقنا هذه الطريقة على مجموعة القيم التي وردت في الجدول الخاص بتنظيم حساب المتوسط للقيم وتكرارها واعتمدنا المتوسط الفرضي (5) لأصبح الوضع كما في الجدول التالي:

الجدول رقم (19): تنظيم حساب المتوسط اعتماداً على متوسط فرضي:

القيم	التكرار (ك)	الانحراف عن المتوسط الفرضي (ح)	المقدار الكلي للانحراف (ح×ك)
10	3	5+	15+
9	2	4+	8+
8	4	3+	12+
7	10	2+	20+
6	8	1+	8+
5	14	صفر.	
4	12	1 –	12 –
3	5	2 –	10 –
2	2	3 –	6 –
1	2	4 –	8 –
	ن = 62		مج (ح×ك) =27

وبتطبيق المعادلة السابقة والتعويض عنها يكون المتوسط هو:

$$5.43 = \frac{27}{62} + 5$$

ويمكن استخدام الطريقة السابقة حين تكون القيم منظمة ضمن فئات بأخذ فئات معينة كمتوسط فرضي، ثم حساب المتوسط على هذا الأساس وذلك وفق المعادلة التالية:

$$م = م\,صفر + \frac{مج\left(حَ \times ك\right) \times ف}{ن}$$

حيث يشير الحرف:

م صفر: إلى المتوسط الفرضي والذي يحسب بأخذ مركز الفئة المعتمدة كمتوسط فرضي.

حَ: إلى انحراف كل فئة مرتبة عن الفئة المأخوذة كمتوسط فرضي.

ف: إلى مدى الفئة.

ولو طبقنا هذه الطريقة على فئات القيم التي وردت سابقاً وافترضنا الفئة المتوسطة الفرضية هي (35 – 39) لأصبح الوضع كما في الجدول التالي (حيث مدى الفئة هو 5).

الجدول رقم (20): تنظيم حساب المتوسط اعتماداً على متوسط فرضي لفئات من القيم:

فئات القيم	التكرار (ك)	الانحراف مرتبة (حَ)	الانحراف × التكرار (حَ×ك)
54 – 50	2	3+	6+
49 – 45	3	2+	6+
44 – 40	5	1+	5+
39 – 35	10	صفر	
34 – 30	6	1 –	6 –
29 – 25	5	2 –	10 –
24 – 20	2	3 –	6 –
المجموع	ن=33		مج حَ ك= – 5

وبتطبيق المعادلة السابقة والتعويض عنها يكون المتوسط هو:

37 (أي مركز الفئة المعتمدة كمتوسط فرضي) + $\frac{5\times5-}{33}$

$$36.24 = \frac{25-}{33} + 37 =$$

وبالطبع يمكن تطبيق هذه الطريقة باستخدام مواضع أخرى للفئة المتوسطة الفرضية، فالمتوسط لا يتأثر بتغير الفئة المأخوذة كفئة متوسطة فرضية، وستكون النتيجة واحدة مهما كانت هذه الفئة.

الوسيط:

إن الوسيط هو أحد مقاييس النزعة المركزية، وهو مقياس هام وشائع الاستعمال. ويعرّف بأنه تلك النقطة من التوزيع لتي يكون عدد القيم المرتبة فوقها مساوياً عدد القيم المرتبة تحتها، أو أنه نقطة المنتصف في سلسلة من القيم.

فمثلاً العلامة 16 هي وسيط العلامات 6، 9، 15، 16، 23، 25، 30 والوسيط للعلامات 17، 22، 33، 44، 46، 50، 53، 59 هو نقطة الوسط بين العلامة 44 والعلامة 46 أي حاصل جمع هاتين العلامتين مقسوماً على 2، أي:

$$45 = \frac{46 + 44}{2}$$

والمعادلة الخاصة بحساب الوسيط لسلسلة من القيم هي:

$$\text{و} = \frac{\text{ن} + 1}{2}$$

حيث يشير الحرف:

و: إلى الوسيط.

ن: إلى عدد القيم

فإذا رجعنا إلى المثال الأول وطبقنا المعادلة يكون:

$$\text{و} = \frac{7 + 1}{2} = 4$$

أي ان الوسيط هنا هو القيمة الرابعة سواء أبدأنا من الأعلى أم بدأنا من الأدنى. وإذا رجعنا إلى المثال الثاني وطبقنا المعادلة يكون:

$$\text{و} = \frac{8 + 1}{2} = 4.5$$

أي ان الوسيط هنا يأخذ مركزاً وسطاً بين القيمتين الرابعة والخامسة وهو كما ذكرنا:

$$45 = \frac{46 + 44}{2}$$

ولكن في حالات معينة قد يكون من الضروري أن يحسب الوسيط من توزيع تكراري مجمع وبخاصة وبخاصة إذا كانت القيم متعددة وكان تكرارها متنوعاً. لنفترض أن لدينا مجموعة من العلامات وقد تم تجميعها في فئات كما في الجدول التالي:

الجدول رقم (21): تنظيم حساب الوسيط لفئات من العلامات (القيم):

فئات العلامات	التكرار	التكرار المجمع
54 – 50	2	33
49 – 45	3	31
44 – 40	5	28
39 – 35	10	23
34 – 30	6	13
29 – 25	5	7
24 – 20	2	2
	ن= 33	

إن المطلوب في هذا المثال هو إيجاد الوسيط، أي النقطة التي تقع تحتها 50٪ من عدد العلامات ومن أجل ذلك يجب اتباع الخطوات التالية:

1. نحدد 50٪ من العلامات بقسمة هذا العدد على 2 وفي مثالنا هذا 50٪ من عدد العلامات= $\frac{33}{2} = 16.5$

2. نحدد في عمود التكرار المجمع قيمة تساوي العدد 16.5 أو تقل عنه مباشرة (وهي هنا القيمة 13). ويقع الوسيط في فئة العلامات التي تعلو مباشرة هذه القيمة من التكرار المجمع وهي الفئة 35 – 39.
3. نطرح قيمة التكرار المجمع التي حصلنا عليها في الخطوة السابقة وهي 13 من نصف عدد العلامات وهي 16.5 والنتيجة هي:

$$3.5 = 13 - 16.5$$

4. نقسم النتيجة السابقة على تكرار الفئة التي سيقع فيها الوسيط ونضرب هذه النتيجة بمدى الفئة أي:

$$1.75 = 5 \times \frac{3.5}{10}$$

5. نضيف إلى النتيجة السابقة الحد الأدنى الحقيقي للفئة التي سيقع فيها الوسيط وذلك بطرح 0.5 من الحد الأدنى لهذه الفئة. والحد الأدنى الحقيقي هنا هو 35 – 0.5 =34.5

وبإضافة القيمة 1.75 إلى 34.5 نحصل على الوسيط وهو 36.25، والمعادلة التالية تلخص الخطوات السابقة جميعاً:

$$\text{و} = \text{حد}_{\text{ف}} + \left(\frac{\frac{\text{ن}}{2} - \text{ك}_{\text{ف}}}{\text{ك}_{\text{ف}}} \right) \text{ف}$$

حيث:

حد ف = الحد الأدنى الحقيقي للفئة التي يقع فيها الوسيط.

ن= العدد.

ك ن = عدد القيم الموجودة تحت الفئة التي يقع فيها الوسيط.

ك ف = تكرار الفئة التي يقع فيها الوسيط.

ف = مدى الفئة.

وبتطبيق هذه المعادلة على المثال الحالي نحصل على ما يلي:

$$و = 34.5 + \left(\frac{\frac{33}{2} - 13}{10} \right) 5$$

$$و = 34.5 + \frac{17.5}{10} = 36.25$$

المنوال:

من الطرائق السهلة للتعرف على القيمة الوسطى أو نقطة التمركز لمجموعة من القيم أو العلامات هي أن نبحث عن القيمة أو العلامة التي تتكرر أكثر من سواها، وهذا هو ما يدعى المنوال أو الشائع. فلو نظرنا في مجموعة العلامات التالية: 12، 14، 14، 14، 15، 17، 17، 18 لوجدنا أن العلامة التي تكررت أكثر من سواها ضمن هذه المجموعة من العلامات هي العلامة 14 وهي بالتالي تمثل المنوال أو الشائع لهذه المجموعة. ولو ألقينا نظرة على جدول التوزيع التكراري السابق لفئات من العلامات لوجدنا أن الفئة 35 – 39 كانت أكثر الفئات تكراراً وهي بالتالي تمثل "الفئة المنوالية" لتلك الفئات من العلامات. وعموماً فإن المنوال: هو وسيلة بسيطة تدلنا بصورة أولية على مركز التجمع للقيم أو العلامات، وهو حساس جداً لأي تغييرات في تلك القيم أو العلامات حتى ولو كانت طفيفة، ففي سلسلة العلامات السابقة لو كانت إحدى العلامات هي 17 بدلاً من 14 لتغيرت قيمة المنوال بصورة ملحوظة وأصبحت العلامة 17 هي المنوال بدلاً من العلامة 14. ولذا يعد المنوال مؤشراً فجاً ومحدود الفائدة وهو دون

المتوسط والوسيط من حيث الدقة، وتنحصر قيمته في أنه يمدنا بفكرة أولية وسريعة عن مركز التجمع أو نقطة التمركز لمجموعة من القيم أو لفئات منها.

استعمالات المتوسط والوسيط والمنوال:

لكل من مقاييس النزعة المركزية التي سبق ذكرها فوائده واستعمالاته الخاصة. فإذا توخى العامل في القياس الدقة التامة، وكان يرغب في حساب قياسات معقدة كالانحراف المعياري ومعامل الارتباط فمن الأفضل استخدام المتوسط بوصفه المقياس الأدق والأقدر على إعطاء أية قيمة في التوزيع وزنها. إلا أن المتوسط يتأثر بشدة بالعلامات المتطرفة، وقد لا يمدنا بصورة صحيحة عن مركز التجمع أو النزعة المركزية لتوزيع توجد فيه علامة واحدة متطرفة أو عدد قليل من العلامات المتطرفة. ونورد المثال التالي توضيحاً لذلك:

لنفترض أن لدينا مجموعة من العلامات مرتبة تصاعدياً على النحو التالي:

16، 18، 21، 22، 22، 23، 24، 25، 26، 28، 58، 62

فالمتوسط في هذه الحالة وهو (28.75) أعلى من (10) من العلامات ولا يقدم صورة صحيحة عن العلامة العادية أو مركز التجمع على حين أن الوسيط لهذه العلامات وهو (23.5) قد يعطينا فكرة أفضل عن العلامة العادية أو نقطة التمركز في التوزيع. وعموماً فإنه حين تكون هناك قيم متطرفة تؤثر في المتوسط بشدة فمن الأفضل استعمال الوسيط. والشيء نفسه يقال حين نرغب في معرفة القيمة الوسطى بسرعة وسهولة ولا نتوخى الدقة التامة. وأما إذا كنا نسعى إلى معرفة القيمة الأكثر تكراراً والحصول على فكرة أولية أو تقدير سريع وتقريبي لنقطة التمركز في التوزيع فيمكن استخدام المنوال أو الشائع.

المئينات ومقاييس التشتت:

تنطوي دراسة تشتت القيم وتباعدها أو انحرافها عن المتوسط على أهمية قصوى نظراً لأن المتوسط لا يعبر بحد ذاته عن انتشار القيم، وقد يكون المتوسط واحداً لمجموعتين من القيم وتكون القيم في واحدة منهما متجمعة حول هذا المتوسط وتكون القيم في الأخرى متباعدة عنه نحو الأدنى والأعلى. لاحظ المجموعتين التاليتين من العلامات:

40، 45، 50، 55، 60

10، 20، 50، 80، 90

إن المتوسط في المجموعة الأولى من العلامات هو (50)، وتنزع هذه العلامات إلى التجمع والاقتراب نحو المتوسط بخلاف المجموعة الثانية من العلامات التي يبتعد أكثرها عن المتوسط بمسافة كبيرة على الرغم من أن متوسطها يعادل متوسط المجموعة الأولى وهو (50). ومقاييس التشتت كثيرة، ومن أكثرها شيوعاً واستخداماً في الدراسات التربوية المقاييس التالية:

المدى:

ويُعبّر عن الفرق بين القيمة الأدنى والقيمة الأعلى في التوزيع، فالعلامات التي تتوزع بين 0 و20 يكون مداها 20 والعلامات التي تتوزع بين 20 و52 يكون مداها 52 – 20=32. وتنحصر فائدة المدى أو المدى المطلق كما يُسمّى أحياناً في أنه يعرفنا على طريق التوزيع، وقد يعطي صورة مضللة إذا كانت بعض القيم متطرفة جداً أو إذا اختلف تكرار بعضها اختلافاً كبيراً عن تكرار بعضها الآخر. وهذا يعني أن المدى يعتمد على العلامتين المتطرفتين بصورة كلية ويمكن تغييره بمجرد إضافة أو حذف قيمة واحدة متطرفة.

الربيع:

يُعرّف الربيع على أنه تلك النقطة في التوزيع التي تقع تحتها ¼ الحالات أو القيم إذا تم ترتيبها بصورة تصاعدية بحيث تبدأ بالأدنى وتنتهي بالأعلى. وهكذا فالربيع الأول يغطي 25٪ من الحالات حين نسير من القيمة الأدنى صعوداً نحو الأعلى (المئين الخامس والعشرون). والربيع الثاني هو تلك النقطة التي تقع تحتها 50٪ من الحالات أي الوسيط نفسه (المئين الخمسون). وأما الربيع الثالث فهو تلك النقطة التي تقع تحتها 75٪ من الحالات (المئين الخامس والسبعون). ومن فوائد الربيع أنه يفيد في معرفة مدى تجمع القيم حول الوسيط من خلال استبعاد القيم المتطرفة التي تقع عادة في الربيع الأول والأخير والإبقاء على الربيعين الثاني والثالث وهو ما يعرف بنصف المدى الربيعي أو مدى ما بين الربيعين.

مدى ما بين الربيعين:

يغطي مدى ما بين الربيعين القيم التي يضمها الربيعان الثاني والثالث أو 50٪ من الحالات الواقعة في وسط التوزيع، وذلك بعد إهمال القيم التي حصل عليها الربع الأعلى والقيم التي حصل عليها الربع الأدنى من الأفراد. ويتسم مدى ما بين الربيعين بأنه أقرب إلى الاعتدال نظراً لاستبعاد القيم المتطرفة نحو الأدنى أو الأعلى، وهو يدلنا على درجة بعد كل من الربيع الثاني والربيع الثالث عن الوسيط الذي يقع في نهاية الربيع الثاني كما ذكرنا. ولمدى ما بين الربيعين فائدته حين يرتئي الباحث إهمال الحالات المتطرفة والاقتصار على دراسة الحالات التي تميل إلى التجمع حول نقطة الوسط في التوزيع فقط.

المئين:

يمثل المئين مقياساً إحصائياً شديد الأهمية نظراً لاستخداماته الواسعة في الاختبارات المعيرة وبساطته وسهولة حسابه. ويمكن النظر إلى المئين على أنه شكل معدل ومحسن للعلامة الخام فحين يتعرف التلميذ على مرتبته المئينية يدرك في الحال مركزه النسبي بين زملائه في المجموعة، ويتضح مستوى أدائه بالمقارنة مع زملائه من خلال معرفه النسبة المئوية للحالات (أو العلامات) التي تقع تحت علامته. فإذا كانت المرتبة المئينية لأحد التلاميذ هي 70 مثلاً فهذا يعني أنه حصل على علامة أعلى من 70٪ من علامات المجموعة. وللتقدير عن طريق المرتبة المئينية أفضليته على التقدير الرتبي الاعتيادي من حيث أن هذا الأخير لا يصلح للمقارنة بين المجموعات المختلفة من حيث العدد، فالخامس في صف مؤلف من 50 طالباً لا تعادل مرتبته مرتبة الخامس في صف مؤلف من ستة طلاب، ولو تم تحويل هذه الرتب إلى رتب مئينية لأصبح من اليسير إجراء المقارنة بدقة ووضوح. وبطبيعة الحال فإن المئين الخامس والعشرين يمثل الربيع الأول على حين أن المئين الخمسين يمثل الربيع الثاني وهو الوسيط، والمئين الخامس والسبعين يمثل الربيع الثالث. ومن المفيد الإشارة هنا إلى أن مصطلحي "المئين" و"المرتبة المئينية" متقاربان جداً. فالمرتبة المئينية تمثل النسبة المئوية للحالات التي تقع تحت علامة معينة في حين أن المئين هو العلامة التي تقع تحتها النسبة المئوية المعطاة " (لندفال، 1968، ص269). وهذا يعني أن المرتبة المئينية التي تعكس أو تقابل علامة خام معينة تشير إلى النسبة المئوية للأشخاص الذين تقع علاماتهم تحت هذه العلامة، فالمرتبة المئينية الثمانون مثلاً لعلامة خام مقدارها 44 تبين أن المئين الثمانين هو 44 وهكذا.

ومن الطرائق المتبعة في حساب المرتبة المئينية للعلامة الخام استخدام المعادلة التالية:

المرتبة المئينية للعلامة الخام = $\frac{\text{عدد العلامات التي تقع تحت هذه العلامة} + \frac{1}{2}\text{تكرارها}}{\text{عدد التلاميذ}} \times 100$

ولنأخذ كمثال التوزيع التكراري التالي لمجموعة من العلامات:

الجدول رقم (22): تنظيم حساب المرتبة المئينية للعلامة الخام:

العلامات (القيم) (س)	التكرار (ك)
10	3
9	2
8	4
7	10
6	8
5	14
4	12
3	5
2	2
1	2
	مج = 62

وبتطبيق المعادلة السابقة على العلامة الخام 8 مثلاً نحصل على المرتبة المئينية التي تقابل هذه العلامة وهي:

$$\frac{2+53}{62} \times 100 = 89$$ تقريباً

أما المرتبة المئينية للعلامة الخام 4 فهي:

$$\frac{6+9}{62} \times 100 = 24$$ تقريباً

الانحراف المتوسط:

يُعرّف الانحراف المتوسط على أنه متوسط مجموع انحرافات القيم عن متوسطها، ويحسب باستخدام المعادلة التالية:

$$\text{حم} = \frac{\text{مج ح}}{\text{ن}}$$

حيث: يشير الرمز:

حم إلى الانحراف المتوسط.

مج ح: إلى مجموع انحرافات القيم عن متوسطها.

ن: إلى عدد القيم.

فإذا كانت لدينا مجموعة من العلامات وأردنا حساب انحرافها عن المتوسط فلابد أن نحسب متوسطها أولاً ثم انحراف كل منها عن المتوسط، ثم نجمع هذه الانحرافات، ونقسم هذا المجموع على العدد كما في المثال التالي:

الجدول رقم (23): تنظيم حساب الانحراف المتوسط:

العلامات	المتوسط	الانحراف عن المتوسط
11	15	4
13	15	2
14	15	1
15	15	0
17	15	2
20	15	5
		مج ح = 14

ويتطبيق المعادلة السابقة يكون الانحراف المتوسط لهذه المجموعة من العلامات هو:

$$\frac{14}{6} = 2.33 \text{ تقريباً}$$

وإذا كانت العلامات مجمعة في فئات ومنظمة في جدول توزيع تكراري فلابد أن نحسب أولاً انحراف مركز كل فئة عن المتوسط، ونضربه بتكرار الفئة، ثم نجمع حواصل الضرب، ونقسمها على عدد القيم أو العلامات، فنحصل على الانحراف المتوسط وذلك وفق المعادلة التالية:

$$\text{حم} = \frac{\text{مج}\left(\text{ك ح}\right)}{\text{ن}}$$

حيث يشير الرمز:

حم: إلى الانحراف المتوسط.

مج (ك ح): إلى مجموع حواصل ضرب انحراف مركز كل فئة عن المتوسط بالتكرار في تلك الفئة.

ن: إلى عدد القيم.

ولنأخذ المثال التالي لفئات من العلامات بلغ متوسطها 36.24

الجدول رقم (24): تنظيم حساب الانحراف المتوسط لفئات من القيم:

فئات العلامات	التكرار	مركز الفئة	الانحراف عن المتوسط	حاصل ضرب الانحراف بالتكرار
54 – 50	2	52	15.76	31.52
49 – 45	3	47	10.76	32.28
44 – 40	5	42	5.76	28.8
39 – 35	10	37	0.76	7.6
34 – 30	6	32	4.24	25.44
29 – 25	5	27	9.24	46.2
24 – 20	2	22	14.24	28.48
المجموع	33			200.32

وبتطبيق المعادلة السابقة وهي:

$$\text{حم} = \frac{\text{مج}\left(\text{ك ح}\right)}{\text{ن}}$$

يكون الانحراف المتوسط لهذه المجموعة من فئات العلامات هو:

$$\frac{200.32}{33} = 6 \text{ تقريباً}$$

الانحراف المعياري:

يعد الانحراف المعياري من أهم مقاييس التشتت وأكثرها استخداماً في البحوث التربوية، وهو يشبه إلى حد كبير الانحراف المتوسط من حيث أنه يعبر عن نوع من متوسط الانحرافات، ولكن يختلف عنه في أنه يتطلب تربيع انحراف كل قيمة عن المتوسط، ثم جذر مجموع الانحرافات بهدف التخلص من الإشارات السالبة التي تصبح موجبة بالتربيع. ويمكن تعريف الانحراف المعياري على أنه

الجذر التربيعي لمجموع مربعات انحرافات القيم عن متوسطها، ومن أبسط طرائق حسابه الطريقة التي تعتمد على استخدام المعادلة التالية:

$$ع = \sqrt{\frac{مج ح^2}{ن}}$$

حيث يشير الحرف: ع إلى الانحراف المعياري

مج $ح^2$: إلى مجموع مربعات انحرافات القيم عن المتوسط.

ن: إلى عدد القيم.

فإذا أردنا حساب الانحراف المعياري لمجموعة من القيم فلابد أولاً أن نحسب متوسطها، ثم انحراف كل منها عن المتوسط، ثم مربع هذا الانحراف، وأخيراً جمع مربعات الانحرافات، وتقسيمها على عدد القيم، وذلك كما في الجدول التالي:

الجدول رقم (25): حساب الانحراف المعياري لمجموعة من العلامات:

العلامات	الانحراف عن المتوسط	مربع الانحراف عن المتوسط
18	8+	64
16	6+	36
15	5+	25
12	2+	4
11	1+	1
7	3 –	9
6	4 –	16
3	7 –	49
2	8 –	64
مج = 90		268

وبتطبيق المعادلة السابقة يكون الانحراف المعياري لهذه المجموعة من العلامات هو:

$$ع = \sqrt{\frac{مج ح^2}{ن}}$$

$$ع = \sqrt{\frac{268}{9}} = 5.45 \text{ تقريباً}$$

ولحساب الانحراف المعياري لتوزيع تكراري للقيم أو العلامات نستخدم المعادلة السابقة بعد إدخال بعض التعديل عليها لتتلاءم مع تكرارات القيم.

وتأخذ هذه المعادلة الشكل التالي:

$$ع = \sqrt{\frac{مج ك ح^2}{ن}}$$

حيث يشير الحرف:

ع: إلى الانحراف المعياري.

مج ك $ح^2$: إلى مجموع حواصل ضرب مربع انحراف كل قيمة مع تكرارها.

ن: إلى عدد القيم.

ويوضح المثال الوارد في الجدول التالي كيفية حساب الانحراف المعياري لتوزيع تكراري للعلامات بهذه الطريقة مع العلم أن المتوسط في هذا التوزيع هو 5.47.

الجدول رقم (26): حساب الانحراف المعياري لتوزيع تكراري:

العلامات	التكرار	الانحراف عن المتوسط	مربع الانحراف عن المتوسط	حاصل ضرب التكرار في مربع الانحراف
10	3	4.53+	20.52	61.56
9	2	3.53+	12.46	24.92
8	4	2.53+	6.40	25.60
7	10	1.53+	2.34	23.40
6	8	0.53+	0.28	2.24
5	14	0.47 –	0.22	3.09
4	12	1.47 –	2.16	25.93
3	5	2.47 –	6.10	30.50
2	2	3.47 –	12.04	24.08
1	2	4.47 –	19.98	39.96
المجموع	ن= 62			مج ك ح^2= 261.28

وبتطبيق المعادلة السابقة وهي:

$$ع = \sqrt{\frac{مج\ ك\ ح^2}{ن}}$$

يكون الانحراف المعياري لهذا التوزيع التكراري للعلامات هو:

$$ع = \sqrt{\frac{261.28}{62}} = \sqrt{4.21} = 2.05$$

ولحساب الانحراف المعياري لقيم كثيرة ومجمعة ضمن فئات في جدول توزيع تكراري قد لا تُجدي الطريقتان السابقتان نظراً لأنهما تتطلبان جهداً كبيراً ووقتاً طويلاً، ويمكن اتباع طريقة أخرى تختلف عن الطرائق السابقة في أنها تقوم على أخذ فئة ما واعتبارها متوسطاً فرضياً ثم حساب انحراف الفئات الأخرى عنها وذلك وفق المعادلة التالية:

$$ع = ف \sqrt{\frac{مج\ ك\ ح^2}{ن} - \left(\frac{مج\ ك\ ح}{ن}\right)^2}$$

حيث يشير الحرف:

ع: إلى الانحراف المعياري.

ف: إلى مدى الفئة.

مج ك ح: إلى مجموع حاصل ضرب تكرار كل فئة في انحرافها مرتبة.

ن: إلى عدد القيم.

ولنأخذ المثال التالي لهذه الطريقة مع ملاحظة أن الفئة المأخوذة كمتوسط فرضي في هذا المثال هي الفئة 35 – 39.

الجدول رقم (27): تنظيم حساب الانحراف المعياري لفئات من القيم اعتماداً على متوسط فرضي:

فئات القيم	التكرار	الانحراف مرتبة	حاصل ضرب التكرار في الانحراف مرتبة	حاصل ضرب التكرار في مربع الانحراف مرتبة
54 – 50	2	3+	6	18
49 – 45	3	2+	6	12
44 – 40	5	1+	17+5	5
39 – 35	10	صفر		
34 – 30	6	1 –	6 –	6
29 – 25	5	2 –	10 –	20
24 – 20	2	3 –	22 – 6 –	18
المجموع	ن=33		مج ك ح= – 5	مج ك $ح^2$= 79

وبالتعويض عن المعادلة السابقة وهي:

$$ع = ف \sqrt{\frac{\text{مج ك ح}^2}{ن} - \left(\frac{\text{مج ك ح}}{ن}\right)^2}$$

يكون الانحراف المعياري لهذه الفئات من القيم هو:

$$ع = 5 \sqrt{\frac{79}{33} - \left(\frac{-5}{33}\right)^2}$$

$$ع = 5\sqrt{2.39 - 0.02} = 1.53$$

وثمة طرائق أخرى عديدة لحساب الانحراف المعياري لا يتسع المقام للوقوف عندها، والمهم في الأمر هو أن الانحراف المعياري من أهم مقاييس التشتت، إن لم يكن أهمها على الإطلاق، وأكثرها دقة وهو يمثل الأساس في حساب العلامة المعيارية وفي الحصول على معامل الترابط (معامل بيرسون) كما سنرى. وليس ثمة تفسير بسيط لمعنى الانحراف المعياري، فمن الصعب أن نقول بكلمات بسيطة ماذا يعني هذا المصطلح. ولكن يمكن القول: إن الانحراف المعياري يعبر عن درجة انتشار القيم أو العلامات. فكلما ازداد انتشار العلامات ازدادت قيمة الانحراف المعياري وكلما تناقص هذا الانتشار انخفضت قيمته. فإذا كان الانحراف المعياري لتحصيل طلاب صف ما في الرياضيات هو 5 والانحراف المعياري لتحصيل طلاب صف آخر في الرياضيات اختبروا بالمقياس نفسه هو 12 مثلاً، فإن هذا يعني أن طلاب الصف الثاني أكثر تبايناً في التحصيل الرياضي من طلاب الصف الأول.

ويتضح معنى الانحراف المعياري في التوزيع السوي بصورة خاصة حيث أن ثلثي الحالات تقريباً 68.2٪ تقع بين النقطتين اللتين تبعدان بمقدار انحراف معياري واحد عن المتوسط في هذا التوزيع (34.1٪ فوق المتوسط و34.1٪ تحت المتوسط) و95٪ من الحالات تقريباً تقع بين النقطتين اللتين تبعدان بمقدار

انحرافين معياريين عن المتوسط، وجميع الحالات تقريباً تقع بين النقطتين اللتين تبعدان بمقدار (3) انحرافات معيارية عن المتوسط وذلك كما في الجدول التالي:

بين – 3، – 2 انحراف معياري عن المتوسط يقع 2.14٪ من الحالات
بين – 2، – 1 انحراف معياري عن المتوسط يقع 13.59٪ من الحالات
بين – 1، صفر انحراف معياري عن المتوسط يقع 34.13٪ من الحالات
بين صفر، +1 انحراف معياري عن المتوسط يقع 34.13٪ من الحالات
بين +1، +2 انحراف معياري عن المتوسط يقع 13.59٪ من الحالات
بين +2، +3انحراف معياري عن المتوسط يقع 2.14 ٪ من الحالات

وهذه العلاقة الثابتة بين وحدات الانحراف المعياري ونسب الحالات في التوزيع السوي تعطي أهمية فائقة للانحراف المعياري. فانطلاقاً منها يمكن تفسير علامة الفرد بعدد معين من الانحرافات المعيارية فوق أو تحت المتوسط والتعبير عنها بتحديد نسبة الحالات التي تقابلها في المجموعة التي ينتمي إليها الفرد. فإذا كانت علامة الفرد تقابل مثلاً (+1) انحراف معياري فوق المتوسط فهذا يعني أنه يتفوق على 84.13٪ من زملائه في المجموعة (50٪ +34.14٪ =84.13٪) وهكذا. وعموماً فإن الانحراف المعياري يمثل الأساس لمقارنة العلامات مباشرة من اختبار لآخر ومن وقت لآخر. فإذا أخذنا في الحسبان أن توزيع العلامات في الاختبارات التحصيلية واختبارات القدرات يقترب في شكله من شكل التوزيع الاعتدالي السوي برزت أهمية الانحراف المعياري بصورة واضحة وجلية.

التباين:

يُعرّف التباين على أنه مربع الانحراف المعياري ويحسب بالمعادلة التالية:

$$ع^2 = \frac{مج\ ح^2}{ن}$$

حيث يشير الرمز ($ع^2$) إلى التباين

مج $ح^2$: إلى مجموع مربعات انحرافات القيم عن متوسطها الحسابي.

ن: إلى عدد القيم.

وليس من الصعب ملاحظة أن الانحراف المعياري هو الجذر التربيعي للتباين. وعموماً فإن التباين كمقياس إحصائي لتشتت القيم وانتشارها يمثل مجالاً علمياً إحصائياً قائماً بذاته، وهو ينطوي على قدر كبير من الأهمية نظراً لأن تحليل التباين أو التحليل الاحصائي يعد أساساً في الكشف عن دلالة الفروق بين المتوسطات وغيرها من المقاييس الإحصائية المختلفة.

العلامة المعيارية:

تعرف العلامة المعيارية أو الموزونة على أنها عدد وحدات الانحراف المعياري التي تبعدها علامة ما عن المتوسط، وهي تحسب بنسبة انحراف علامة الفرد عن المتوسط إلى انحراف علامات أفراد المجموعة كلها عن هذا المتوسط، وذلك باستخدام المعادلة التالية:

$$ذ = \frac{خ - م}{ع}$$

حيث يشير الحرف (ذ) إلى العلامة المعيارية (الذالية)

(خ) إلى العلامة الخام للفرد

(م) إلى المتوسط

(ع) إلى الانحراف المعياري

فإذا كانت العلامة الخام التي حصل عليها أحد التلاميذ هي (28) وكان متوسط العلامات في المجموعة هو (20) والانحراف المعياري لها هو (4) فإن العلامة المعيارية التي تقابل العلامة الخام لهذا التلميذ هي:

$$2 = \frac{20 - 28}{4}$$

وهذا يعني أن العلامة الخام (28) تبعد بمقدار انحرافين معياريين فوق المتوسط. وأما العلامة المعيارية التي تقابل العلامة الخام (16) التي حصل عليها تلميذ آخر ينتمي إلى تلك المجموعة نفسها فهي:

$$1- = \frac{20 - 16}{4}$$

وهذا يعني أن علامة هذا التلميذ تبعد بمقدار انحراف معياري واحد تحت المتوسط. ولو افترضنا أن هذا التلميذ حصل في اختبار ثان على علامة خام مقدارها 45 وأن متوسط الدرجات لهذا الاختبار هو 50 والانحراف المعياري هو 5 فإن العلامة المعيارية التي تقابل العلامة الخام 45 لهذا التلميذ هي:

$$1- = \frac{50 - 45}{5}$$

وهذا يعني أن علامة هذا التلميذ في الاختبار الثاني ومقدارها 45 تبعد أيضاً بمقدار انحراف معياري واحد تحت المتوسط، وهي بالتالي تعادل علامته في الاختبار الأول وهي 16 وذلك بالنسبة للمتوسط والانحراف المعياري.

وتظهر في هذا المثال أهمية العلامة المعيارية من حيث أنها تمثل أساساً معيارياً للمقارنة يمكن استناداً إليه تفسير العلامات الخام ومقارنة كل منها بمتوسط الأداء في المجموعة من خلال نسبة انحرافها إلى الانحراف المعياري.

وثمة أنواع عديدة من العلامات المعيارية، يطلق عليها اسم العلامات المعيارية المعدلة أو المشتقة. ومنها العلامة (ت) وتشير إلى الحرف الأول من اسم ثورندايك واضعها. وتقوم العلامة (ت) على تحويل العلامات المعيارية (ذ) إلى علامات معيارية اعتدالية متوسطها 50 وانحرافها المعياري 10، وذلك بهدف التخلص من كسور العلامات المعيارية والإشارات السالبة، وتحسب باستخدام المعادلة التالية:

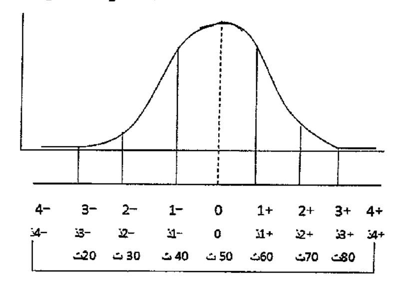
العلامة (ت) = العلامة المعيارية الذاتية (ذ) × 10 +50

فإذا حصل أحد التلاميذ على العلامة الخام (20) وكان متوسط العلامات في المجموعة هو 15 والانحراف المعياري هو 5، فإن علامته المعيارية (ذ) التي تقابل علامته الخام (20) هي:

$$1 = \frac{15 - 20}{5}$$

وأما علامته التائية فهي: 1×10+50 = 60

ويشير الشكل التالي إلى توزيع الدرجات المعيارية (الذالية) ومقابلاتها التائية في منحنى التوزع الاعتدالي أو السوي.



الشكل (13): توزع الدرجات المعيارية (الذائية) ومقابلاتها التائية في المنحني السوي

ومن الجدير بالإشارة هنا أنه في اختبار التصنيف العام للجيش الأمريكي تم حساب الدرجة المعيارية المعدلة بالاستناد إلى متوسط مقداره 100 وانحراف معياري مقداره 20 أي بضرب (ذ) × 20+100 وفي بعض الجامعات الأمريكية يتم حساب الدرجات المعيارية بالاستناد إلى متوسط مقداره 500 وانحراف معياري مقداره 100 أي بضرب (ذ) ×100 +50 وذلك بهدف الكشف عن مستويات الفروق الطفيفة جداً.

ومن الدرجات المعيارية المعدلة الدرجة الجيمية (ج)، وتشير إلى الحرف الأول من اسم جيلفورد. وهي ترتكز على توزيع متوسطه 5 وانحرافه المعياري 2 أي بضرب (ذ)× 2+5، فإذا كانت الدرجة المعيارية (ذ) للفرد هي 0.5 فإن درجته الجيمية المقابلة لهذه الدرجة هي: 0.5 ×2+5 = 6

وقد قام سلاح الطيران الأمريكي بتعديل الدرجة (ج) فبدلاً من تقسيم التوزيع إلى 11 قسماً أصبح يقسمه إلى 9 أقسام، واشتق بذلك معيار التسيع (معيار ستاناين) وقد تم ذلك بضم الدرجة (ج) صفر و(ج)1 في التسيع 1 والدرجة ج (9) والدرجة ج (10) في التسيع 9. ويصلح المعيار الجيمي ومعيار التسيع لكشف

الفروق الكبيرة بين الأفراد نظراً لأنهما يقلصان الفروق إلى عدد ضئيل من المستويات (11 مستوى في الأول و9 مستويات في الثاني).

وتقابل الدرجات التساعية نسب مئوية محددة تحت المنحنى الاعتدالي تظهر على النحو التالي:

التساعي أو التسيع الأول يقابل 4٪ من الحالات.

التسيع الثاني يقابل 7٪ من الحالات.

التسيع الثالث يقابل 12٪ من الحالات.

التسيع الرابع يقابل 17٪ من الحالات.

التسيع الخامس يقابل 20٪ من الحالات.

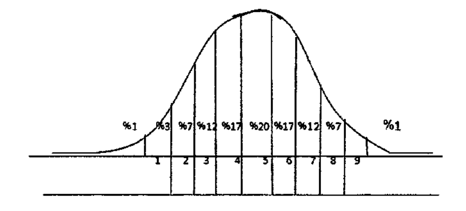
التسيع السادس يقابل 17٪ من الحالات.

التسيع السابع يقابل 12٪ من الحالات.

التسيع الثامن يقابل 7٪ من الحالات.

التسيع التاسع يقابل 4٪ من الحالات.

وتظهر في الشكل التالي الدرجات التساعية وما يقابلها من نسب مئوية تحت المنحني الاعتدالي.



الشكل 14: لدرجات التساعية وما يقابلها من نسب مئوية تحت المنحني الاعتدالي

مقاييس الارتباط:

كثيراً ما يهتم العامل في القياس بمعرفة العلاقة بين متغيرين أو أكثر، والكشف عن قوة واتجاه هذه العلاقة، ويستخدم من أجل ذلك معامل الارتباط. فإذا طبق اختباران على مجموعة من الطلاب أحدهما في الفيزياء والثاني في الكيمياء فإلى أي درجة يمكن القول: إن الطلبة الذين كان أداؤهم جيداً في اختبار الفيزياء كان أيضاً جيداً في اختبار الكيمياء، أو إن الطلبة الذين كان أداؤهم ضعيفاً في اختبار الفيزياء كان أيضاً ضعيفاً في اختبار الكيمياء. ويأخذ معامل الارتباط قيماً تتراوح بين +1 و − 1 مروراً بالصفر. فمعامل ارتباط مقداره (+1) يشير إلى أن الشخص الذي حاز على المرتبة الأولى في الاختبار الأول حاز على المرتبة الأولى في الاختبار الثاني، وأن الذي حاز على المرتبة الثانية في الاختبار الأول حاز على المرتبة الثانية في الاختبار الثاني، وهكذا بالنسبة لكل أفراد المجموعة. ومعامل ارتباط مقداره (− 1) يدل على أن العلامات تسير في اتجاه معاكس تماماً، أي أن الشخص الذي حصل على المرتبة الأولى في الاختبار الأول حصل على المرتبة الأخيرة في الاختبار الثاني، والشخص الذي حصل على المرتبة الثانية في الاختبار الأول كان ترتيبه الثاني قبل الأخير في الاختبار الثاني وهكذا. أما معامل الارتباط الذي مقداره صفر فيشير إلى عدم وجود أي ارتباط بين مجموعتي العلامات. وهذا يعني بعبارة أخرى: ان الارتباط الموجب والتام (+1)

يشير إلى أن الزيادة في المتغير الأول تصحبها زيادة مماثلة في المتغير الثاني، وأن النقصان في المتغير الأول يصحبه نقصان مماثل في المتغير الثاني. وأما الارتباط السالب والتام (− 1) فيشير إلى أن الزيادة في المتغير الأول يصاحبها نقصان مماثل في المتغير الثاني، وأن النقصان في المتغير الأول تصاحبه زيادة مماثلة في المتغير الثاني. على حين أن المعامل صفر يشير إلى انعدام الارتباط. وبطبيعة الحال فإن القيم الواقعة بين (صفر) و(+1) وصفر و(− 1) تدل على وجود ارتباط من درجة ما أو ارتباط جزئي موجب من (صفر) إلى (+1)، وسالب من (صفر) إلى (− 1). ومن الجدير بالإشارة أنه يندر في ميدان التربية أن يصل معامل الارتباط إلى +1 أو −1، والغالب هو ظهور ترابط جزئي إيجابي كالترابط بين تحصيل الأبناء والمستوى الاجتماعي والاقتصادي للأسرة، أوالترابط بين التحصيل في الرياضيات والتحصيل في العلوم.

إن وجود ترابط بين ظاهرتين يعني أن التغير في إحداهما يصحبه تغير في الأخرى، ولا يعني بالضرورة أن هناك علاقة سببية بينهما. فإذا كان هناك ارتباط بين الظاهرة (أ) والظاهرة (ب) فليس معنى هذا بالضرورة أن (أ) سبب لـ (ب) أو أن (ب) سبب لـ (أ)، وقد يكون هناك عامل أو سبب ما يؤثر في كل من (أ) و(ب) ويؤدي إلى ظهور ارتباط من درجة ما بينهما. فظهور ارتباط عالٍ وموجب بين التفوق في الفيزياء والتفوق في الكيمياء مثلاً، لا يعني أن التفوق في الفيزياء هو سبب للتفوق في الكيمياء، أو أن التفوق في الكيمياء هو سبب للتفوق في الفيزياء بل يعني أن أكثر التلاميذ الذين تفوقوا في الفيزياء تفوقوا أيضاً في الكيمياء لأسباب ما لا يمكن الكشف عنها عن طريق دراسة الارتباط، ولابد من البحث عن وسائل أخرى للكشف عنها.

ولدراسة الارتباط وتحديد قيمته أهمية قصوى في القياس النفسي والتربوي. فعن طريق معاملات الارتباط يمكن دراسة خصائص الأسئلة (البنود) وتحديد درجة دقة أداة القياس وثباتها، من خلال الكشف عن درجة الاتساق (التجانس) الداخلي لها. وتستخدم معاملات الارتباط للتحقق من الصدق

التنبؤي لأداة القياس وفي التحديد الكمي للارتباط بين المتغيرات المختلفة كالقدرات والميول وغيرها، هذا بالإضافة إلى استخدامها الواسع في التحليل العاملي للقدرات.

بيد أن الارتباط وإن كان يفيد، كما هو واضح، في دراسة الكثير من متغيرات السلوك الإنساني عن طريق الكشف عن العلاقة الارتباطية بينها فإنه لا يشير إلى علاقة سببية بين متغير وآخر كما أسلفنا، وكل ما يعنيه هو إمكان حدوث عامل عند حدوث الآخر (أو اقتران عامل بآخر) في حالة العلاقة الإيجابية أو السلبية، وإلى انعدام هذا الإمكان حين يصل معامل الارتباط إلى الصفر. وليس ثمة قيمة عددية محددة لمعامل ارتباط يمكن عدها دليلاً على قوة الارتباط بين المتغيرين (أو ضعفه)، إذ يعتمد ذلك إلى حد بعيد على الغرض من حساب هذا المعامل كأن يكون معامل ثبات أو معامل صدق تنبؤي أو غير ذلك. ولكن، بطبيعة الحال، كلما كان معامل الارتباط بين المتغيرين الخاضعين للدراسة أعلى كان أحد المتغيرين أكثر تمشياً مع الآخر واقتراناً به. وثمة طرائق عديدة لحساب معامل الارتباط منها طريقة الرتب لسبيرمان وطريقة الجداء لبيرسون.

حساب معامل الارتباط بطريقة الرتب:

والمعادلة الخاصة بحساب معامل الارتباط بهذه الطريقة هي التالية:

$$ر = 1 - \frac{6 \text{ مج فر}^2}{\text{ن}(\text{ن}^2 - 1)}$$

حيث يشير الحرف (ر) إلى معامل الارتباط

(فر) إلى الفرق بين رتبة الفرد في الاختبار الأول ورتبته في الاختبار الثاني.

ن إلى عدد الأفراد.

والمثال التالي يوضح حساب معامل الارتباط بهذه الطريقة:

حصلت مجموعة من الطلاب على العلامات التالية في اختبارين (كما هو موضح في الجدول التالي):

الجدول رقم (28): حساب معامل الارتباط بطريقة الرتب:

الطلاب	العلامات في الاختبار الأول	العلامات في الاختبار الثاني	الرتبة في الاختبار الأول	الرتبة في الاختبار الثاني	فرق الرتبة	مربع الفرق
أ	6	7	8	8	صفر	صفر
ب	7	9	7	5.5	1.5	2.25
ج	8	12	6	4	2	4
د	10	9	5	5.5	0.5	0.25
هـ	12	8	4	7	3	9
و	13	14	3	2	1	1
ز	15	13	2	3	1	1
ح	18	19	1	1	صفر	صفر
المجموع (8)						17.5

وبالتعويض عن المعادلة السابقة يكون معامل الارتباط بين علامات الاختبار الأول وعلامات الاختبار الثاني هو:

$$ر = 1 - \frac{17.5 \times 6}{8\ (64 - 1)} = 1 - 0.21 = 0.79 \text{ تقريباً}$$

ولابد من ملاحظة أن هناك تلميذين حصلا على علامة واحدة في الاختبار الثاني وهي العلامة 9 ويجب أن يحتلا المرتبتين الخامسة والسادسة.

والشيء الذي نفعله في هذه الحالة للحصول على رتبة كل منهما، هو أن نجمع رتبته مع رتبة زميله الذي حصل على العلامة نفسها، ثم نقسم مجموع الرتبتين وهو 5 + 6 على العدد وهو (2)، فتكون رتبة كل منهما هي: 5.5

حساب معامل الارتباط بطريقة الجداء:

يحسب معامل الارتباط بطريقة الجداء بعدة أشكال يقوم بعضها على حساب الدرجات المعيارية وبعضها الآخر على حساب الانحرافات المعيارية أو الانحرافات فقط، هذا بالإضافة إلى الطريقة المعروفة بالطريقة العامة والتي تعتمد مباشرة على الدرجات الخام ومربعات هذه الدرجات دون الحاجة إلى حساب الدرجات المعيارية أو الانحرافات المعيارية أو الانحرافات. وتستخدم في حساب معامل الارتباط بطريقة الدرجات المعيارية المعادلة التالية:

$$ر = \frac{مج (ذ س \times ذ ص)}{ن}$$

حيث يشير الحرف (ر): إلى معامل الارتباط

ذ س: إلى أية درجة معيارية من درجات الاختبار الأول

ذ ص: إلى الدرجة المعيارية في الاختبار الثاني التي تقابل الدرجة (ذ س)

ن: إلى عدد الأفراد

وتتطلب هذه الطريقة الكثير من المعلومات الحسابية. لذا كثيراً ما يتم استخدام الطريقة الثانية والأبسط والتي لا تتطلب حساب الدرجات المعيارية بل تتطلب حساب الانحرافات المعيارية فقط. والمعادلة الخاصة بحساب معامل الارتباط بطريقة الانحرافات المعيارية هي:

$$ر = \frac{مج (ح س \times ح ص)}{ن ع س ع ص}$$

حيث يشير الرمز (ر): إلى معامل الارتباط.

ح س: إلى انحراف كل قيمة في سلسلة القيم س عن متوسطها.

ح ص: إلى انحراف كل قيمة في سلسلة القيم ص عن متوسطها.

ن: إلى العدد.

ع س: إلى الانحراف المعياري للقيم في (س).

ع ص: إلى الانحراف المعياري للقيم في (ص).

ويوضح الجدول التالي كيفية حساب معامل الارتباط بطريقة الانحرافات المعيارية:

الجدول رقم (29): تنظيم حساب معامل الارتباط بطريقة الانحرافات المعيارية:

الطلاب	العلامات في الاختبار الأول (س)	انحراف العلامات (ح س)	العلامات في الاختبار الثاني (ص)	انحراف العلامات (ح ص)	حاصل ضرب الانحرفات (ح س × ح ص)
أ	2	3 –	3	3 –	9
ب	3	2 –	3	3 –	6
ج	5	صفر	6	صفر	صفر
د	7	2+	8	2+	4
هـ	8	3+	10	4+	12
مج س=25 م س= 5 ع س= 2.28			مج ص=30 م ص=6 ع ص=2.76		مج (ح س× ح ص)= 31

وبالتعويض عن المعادلة السابقة يكون الارتباط المحسوب بهذه الطريقة:

$$\text{ر} = \frac{\text{مج (ح س × ح ص)}}{\text{ن ع س ع ص}}$$

$$\text{ر} = \frac{31}{2.76 \times 2.28 \times 5} = 0.98 \ \text{تقريباً}$$

وأما طريقة حساب معامل الارتباط اعتماداً على الانحرافات فتقوم على الاستغناء عن حساب الانحراف المعياري والاكتفاء بحساب الانحرافات عن المتوسط ومربعاتها، وذلك وفق المعادلة التالية:

$$\text{ر} = \frac{\text{مج (ح س × ح ص)}}{\sqrt{\text{مج ح}^2\text{س × مج ح}^2\text{ص}}}$$

ويوضح الجدول التالي كيفية حساب معامل الارتباط بهذه الطريقة. وسنأخذ درجات المثال السابق نفسه لتسهيل هذه المهمة.

الجدول رقم (30): حساب معامل الارتباط بطريقة الانحرافات:

الطلاب	العلامات في الاختبار الأول (س)	انحراف العلامات (ح س)	مربعات الانحرافات ح2 س	العلامات في الاختبار الثاني (ص)	انحراف العلامات (ح ص)	مربعات الانحرافات ح2 ص	حاصل ضرب الانحرفات (ح س × ح ص)
أ	2	3 –	9	3	3 –	9	9
ب	3	2 –	4	3	3 –	9	6
ج	5	صفر	صفر	6	صفر	صفر	صفر
د	7	2+	4	8	2+	4	4
هـ	8	3+	9	10	4+	16	12
ن= 5	مج س=25 م س=5		مج ح2س=26	مج ص=30 م ص=6		مج ح2 ص=38	مج (ح س× ح ص)=31

وبالتعويض عن المعادلة السابقة يكون معامل الارتباط المحسوب بطريقة الانحرافات ومربعاتها هو:

$$ر = \frac{\text{مج}\left(ح\,س \times ح\,ص\right)}{\sqrt{\text{مج}\,ح^{2}\,س \times \text{مج}\,ح^{2}\,ص}} = \frac{31}{\sqrt{38 \times 26}} = 0.98 \text{ تقريباً}$$

والطريقة العامة في حساب معامل الارتباط تعتمد مباشرة على الدرجات الخام ومربعات هذه الدرجات، وتتميز هذه الطريقة بالسرعة والدقة.

والمعادلة الخاصة بها هي:

$$ر = \frac{\text{ن مج س ص} - \text{مج س} \times \text{مج ص}}{\sqrt{\left(\text{ن مج س}^{2} - (\text{مج س})^{2}\right)\left(\text{ن مج ص}^{2} - (\text{مج ص})^{2}\right)}}$$

حيث يشير الرمز (مج س ص) إلى مجموع حاصل ضرب الدرجات المتقابلة في الاختبارين.

و (مج س× مج ص) إلى حاصل ضرب مجموع درجات الاختبار الأول س في مجموع درجات الاختبار الثاني ص.

و (مج $س^{2}$) إلى مجموع مربعات الاختبار الأول س.

و $(\text{مج س})^{2}$ إلى مربع مجموع درجات الاختبار الأول س.

و (مج $ص^{2}$) إلى مجموع مربعات الاختبار الثاني ص.

و $(\text{مج ص})^{2}$ إلى مربع مجموع درجات الاختبار الثاني ص.

ويوضح الجدول التالي كيفية حساب معامل الارتباط من الدرجات الخام مباشرة (الطريقة العامة) وسنستعين بدرجات المثال السابق هنا أيضاً لتسهيل هذه المهمة.

الجدول رقم (31): حساب معامل الارتباط بالطريقة العامة:

الطلاب	درجات الاختبار الأول (س)	مربعات درجات الاختبار الأول ($س^2$)	درجات الاختبار الثاني (ص)	مربعات درجات الاختبار الثاني ($ص^2$)	حاصل ضرب الدرجات المقابلة س×ص
أ	2	4	3	9	6
ب	3	9	3	9	9
ج	5	25	6	36	30
د	7	49	8	64	56
هـ	8	64	10	100	80
ن= 5	مج س=25 $(مج س)^2$=625	مج $س^2$=151	مج ص=30 $(مج ص)^2$= 900	مج ص2=218	مج س×ص = 181

وبالتعويض عن المعادلة السابقة يكون معامل الارتباط المحسوب بهذه الطريقة هو:

$$ر = \frac{ن\ مج\ س\ ص - مج\ س \times مج\ ص}{\sqrt{((ن\ مج\ س^2 - (مج\ س)^2)(ن\ مج\ ص^2 - (مج\ ص)^2))}}$$

$$ر = \frac{30 \times 25 - 181 \times 5}{\sqrt{(900 - 218 \times 5)(625 - 151 \times 5)}}$$

$$ر = \frac{750 - 905}{\sqrt{(900 - 1090)(625 - 755)}}$$

$$ر = \frac{155}{\sqrt{190 \times 130}} = 0.98 \text{ تقريباً}$$

وبهذه الطريقة تكون القيمة العددية لمعامل الارتباط هي 0.98 تقريباً، وهي القيمة نفسها التي حصلنا عليها بطريقة الانحرافات المعيارية وبطريقة الانحرافات مما يدل على أن جميع الطرائق التي تعتمد على الجداء تؤدي إلى النتيجة نفسها بعد تقريبها.

التحليل الإحصائي لبنود الاختبار:

ينطوي التحليل الإحصائي لبنود الاختبار على أهمية قصوى سواء للباحث أم للمعلم الذي يسعى إلى تأسيس صدق وثبات الاختبار الذي أعدّه. وسنتعرض فيما يلي لبعض الأساليب والإجراءات المتبعة في هذا التحليل والتي تفيد المعلم خاصة ولا تتطلب سوى القيام ببعض العمليات الإحصائية البسيطة.

حساب معاملات السهولة:

يمثل حساب معامل السهولة لكل بند من البنود التي يضمها الاختبار الخطوة الأولى في عملية تحليل البنود. ولحساب هذا المعامل لابد من إعداد جدول بالإجابات الخاصة بكل بند في الاختبار وذلك كما في الجدول التالي (السيد ومخائيل، 1989، ص20).

الجدول رقم (32): نموذج تفريغ الإجابات الصحيحة:

رقم السؤال	عدد الإجابات الصحيحة	عدد الطلبة المفحوصين	النسبة المئوية للإجابات الصحيحة
1			
2			
3			
4			
5			
الخ			

ويكون معامل سهولة السؤال هو النسبة المئوية للإجابات الصحيحة عن هذا السؤال، ويحسب بالصيغة التالية:

$$\text{معامل سهولة السؤال} = \frac{\text{عدد الذين أجابوا إجابة صحيحة عن السؤال}}{\text{عدد من حاول الإجابة عليه من المفحوصين}} \times 100$$

وللحصول على معامل الصعوبة يطرح معامل السهولة من واحد صحيح. فإذا كان معامل سهولة السؤال هو 0.40 فإن معامل صعوبته هو:

$$0.60 = 0.40 - 1$$

ومن المعلوم أن السؤال شديد السهولة أو شديد الصعوبة لا يميز بين المفحوصين، وأفضل الأسئلة من حيث القدرة التمييزية هي الأسئلة التي يصل معامل سهولتها إلى 0.50. وعموماً فإنه من الضروري لرفع القدرة التمييزية للبنود تجنب استعمال البنود ذات مستويات الصعوبة المتباينة تبايناً واسعاً، ويفضل الاقتصار على البنود التي تتراوح معاملات سهولتها ما بين 0.40 و0.60

ويمكن بغرض رفع مستوى الدافعية لدى المفحوصين الاحتفاظ ببعض ابنود السهلة التي يمكن أن يبدأ بهذا الاختبار.

معامل التمييز:

تشير القدرة التمييزية للبند الاختباري إلى قدرته على التفريق أو التمييز بين المفحوصين الذين يظهرون أداءً حسناً والمفحوصين الذين يظهرون أداءً ضعيفاً في الاختبار ككل. وبعبارة أخرى يُعدّ البند مميزاً جيداً إذا ترابط الأداء على هذا البند ترابطاً موجباً وعالياً مع الأداء في الاختبار ككل. والبنود التي لا تميز بين المفحوصين أو تترابط ترابطاً سلبياً مع الأداء الاختباري ككل تقلل التباين في توزيع درجات الاختبار الكلي وتضعف بالتالي مستوى الصدق والثبات.

وثمة مؤشران للقدرة التمييزية: وهما معامل ارتباط بيرسون الذي يُستعمل لقياس درجة الترابط بين أداء الطالب على البند وادائه على الاختبار ككل. والمؤشر الثاني: وهو الأبسط والأسهل في حسابه، ويمكن استعماله بسهولة من قبل معلم الصف، ويُحسب وفق الإجراء التالي:

1. رتّب اوراق الإجابة بدءاً بالعلامة الأعلى في الاختبار الكلي وانتهاءً بالعلامة الأدنى، ثم صنّف هذه الأوراق إلى ثلاث فئات بحيث تأخذ الفئة العليا نسبة 25٪ والوسطى 50٪ والدنيا 25٪. ويمكن زيادة نسبة كل من الفئة العليا والدنيا إلى 27٪ أو 30٪ أو 33.3٪ مع أن هذا غير ضروري وبخاصة في الاختبارات الصفية.
2. اوجد نسبة الأشخاص في الفئة العليا الذين اعطوا إجابات صحيحة عن البند. وهذه النسبة تُشير ببساطة إلى مستوى او معامل السهولة في الفئة العليا فقط.

3. أوجد نسبة الأشخاص في الفئة الدنيا الذين أعطوا إجابات صحيحة عن البند. وهذه النسبة تُشير ببساطة إلى مستوى أو معامل السهولة في الفئة الدنيا فقط.
4. أوجد معامل التمييز أو مؤشر القدرة التمييزية للبند ويُحسب هكذا:

معامل التمييز = معامل السهولة العلوي – معامل السهولة السفلي

فإذا كانت نسبة الإجابات الصحيحة عن البند أو السؤال في الفئة العليا هي 80٪ (معامل السهولة العلوي)، ونسبة الإجابات الصحيحة عن هذا البند أو السؤال في الفئة الدنيا هي 30٪ (معامل السهولة السفلي)، فإن معامل التمييز لهذا السؤال هو 0.80 – 0.30= 0.50 وهذا المعامل يُشير إلى أن السؤال يميز جيداً بين الأقوياء والضعاف. وعموماً فإنه إذا كان معامل التمييز موجباً فهذا يعني أن البند يميز بدرجة ما ويعمل بالاتجاه نفسه الذي يعمل به الاختبار ككل. وإذا كان سالباً فهذا يُشير إلى العكس، أي أن الأشخاص الذين حصلوا على درجات منخفضة في الاختبار الكلي يميلون إلى إعطاء إجابات صحيحة عن البند بينما الأشخاص الذين حصلوا على درجات عالية يميلون إلى إعطاء إجابات غير صحيحة عن البند. وهذا ما يُشير إلى وجود خلل في البند، أي أن البند إما أن يُساء تفسيره من قبل الأقوياء أو أنه يقدم تلميحاً للضعاف. وإذا اقتربت قيمة معامل التمييز من الصفر، فهذا يعني أن البند يفتقر إلى القدرة التمييزية. وعموماً فإن البنود التي يقل معامل التمييز فيها عن 0.20 لا تُعدّ مرغوبة ويُستحسن تعديلها أو حذفها. وينصح المعلم بإعداد جدول لتفريغ الإجابات من أجل حساب معامل التمييز وذلك وفق النموذج التالي: (السيد ومخائيل، 1989، ص21).

الجدول رقم (33): نموذج تفريغ الإجابات لحساب معامل التمييز:

رقم السؤال	عدد الإجابات الصحيحة في الفئة العليا (الثلث الأعلى)	النسبة المئوية	عدد الإجابات الصحيحة في الفئة الدنيا (الثلث الأدنى)	النسبة المئوية
1				
2				
3				
4				
5				
الخ				

دراسة فعالية المشتتات (المموّهات):

تتطلب عملية تحليل البنود بالإضافة إلى حساب معاملات السهولة والتمييز، دراسة فعالية المشتتات (أو المموهات) في أسئلة الاختيار من متعدد للتأكد مما إذا كانت تقوم بالدور المسند إليها وهو" تشتيت" انتباه الطلبة غير العارفين ومنعهم من الوصول إلى الجواب الصحيح بمحض المصادفة. ويمكن تحديد الأغراض الخاصة لدراسة المشتتات فيما يلي:

1. التأكد مما إذا كانت جذابة ومغرية للطلبة بحيث يختارها بعضهم أم أنها غير جذابة لدرجة أن حذفها أو إبقاءها سيان.
2. التأكد مما إذا كان عدد الذين تجذبهم في الفئة الدنيا أكبر منه في الفئة العليا. ومن الواضح أن المشتت إذا لم يجلب أحداً أو جذب عدداً ضئيلاً للغاية، أو كان عدد من جذبهم في الفئة العليا أكبر منه في الفئة الدنيا، فإنه يعدّ غير فعال ولابد من تعديله أو حذفه ليحل محله مشتت آخر يمكن أن يؤدي دوره على النحو المطلوب. وهاك مثالاً لإجابات مئة من المفحوصين

على بند اختباري من نوع الاختيار من متعدد يحتوي على أربعة بدائل بينها البديل (ب) هو الصحيح والبدائل الأخرى (أ) و(ج) و(د) هي مموّهات:

الجدول رقم (34): صحيفة الإجابات على بند اختياري متعدد الاختيار مؤلف من أربعة بدائل:

الفئات	البدائل				المجموع
	أ	ب	ج	د	
الفئة العليا 25٪	3	18	3	1	25
الفئة الوسطى 50٪	15	20	13	2	50
الفئة الدنيا 25٪	8	10	7	0	25
المجموع	26	48	23	3	100

وبالقاء نظرة سريعة إلى هذا الجدول يتبين أن البند يقترب إلى مستوى المتوسط من حيث السهولة، حيث يبلغ معامل سهولته 0.48، كما يتبين ان البند يتمتع بقدرة تمييزية لا بأس بها حيث يبلغ معامل تمييزه 0.72 – 0.40 = 0.32 وأن المموّه (أ) يمكن عده مناسباً وفعالاً حيث كان عدد من اختاره من الفئة الدنيا أكبر بشكل ملحوظ من عدد من اختاره من الفئة العليا (8 إلى 3). وكذلك بالنسبة للمموّه (ج) فقد اختار سبعة من المفحوصين من الفئة الدنيا مقابل ثلاثة من الفئة العليا، وأما المموّه (د) فلا يمكن عدّه مناسباً حيث لم يتم اختياره من قبل أي من المفحوصين من الفئة الدنيا واختاره أحد المفحوصين من الفئة العليا. وهذا المموّه الأخير لابد من إقصائه أو تعديله بإعادة صياغته من جديد.

العينات وطرائق اختيارها:

يشغل موضوع العينات حيزاً هاماً في البحث التربوي والقياس نظراً لأن البحوث التربوية تعتمد في أغلب الحالات، إن لم يكن في جميع الحالات، على اختيار عينة بطريقة ما، ولأن المقياس في المجالات النفسية والتربوية هو بحد ذاته عينة من البنود أو الأسئلة يفترض أن تمثل الإجابة عنها الإجابة عن جميع البنود أو الأسئلة المحتملة التي يضمها الموضوع الخاضع للقياس.

وتنبع أهمية موضوع العينات من أنه يدخل مباشرة في نطاق الاستدلال الإحصائي ويقوم على استخلاص " الخواص الإحصائية للأصل من الخواص الإحصائية لإحدى أو بعض عيناته، أي أنه يستنتج صفات الكل من الجزء أو الأجزاء التي تنطوي تحت إطاره" (البهي السيد، 1978، ص304)

ومما يظهر أهمية العينات في التربية وعلم النفس أننا حين نرغب في إجراء دراسة حول ظاهرة نفسية أو تربوية ما على عدد كبير من الأفراد نحتاج إلى وقت وجهد كبيرين، كما نحتاج إلى نفقات مادية كبيرة، مما قد يؤدي إلى تعذر إجراء هذه الدراسة أو استحالتها وبخاصة إذا كان عدد الأفراد موضع الدراسة كبيراً للغاية. ويمكن في مثل هذه الحالات أن نقتصر على اختيار عينة من الأفراد شريطة أن تتمثل فيها جميع صفات الأصل الذي اشتقت منه ليكون استنتاجنا صحيحاً حول هذا الأصل أو المجتمع الأصلي. ومن الواضح أن الحكمة من وراء اختيار العينة هي الاقتصاد في الوقت والجهد والنفقات المادية وأن هذا النزوع إلى الاقتصاد يجب ألا يكون على حساب صدق العينة وتمثيلها للمجتمع الأصلي الذي اشتقت منه بحيث تأتي صورة صادقة لهذا المجتمع بجميع صفاته وملامحه، أو بصفاته وملامحه الهامة على الأقل. وتتوقف على العينة بالذات كل القياسات والنتائج التي تسفر عنها عملية البحث. فإذا لم تكن العينة صادقة بدرجة كافية انعكس ذلك على البحث ونتائجه بصورة مباشرة مما يؤدي بطبيعة الحال إلى التشكيك بهذا البحث وقيمته وجدوى الجهود التي بذلت فيه.

المسائل التي تثيرها العينات:

مما لا يرقى إليه الشك أن استخدام أسلوب العينات ينطوي على قدر كبير من الأهمية نظراً لفعالية هذا الأسلوب في اختصار الوقت والجهد والمال إلى درجة قد تكون هائلة في الكثير من الحالات. إلا أن لأسلوب العينة، مع ذلك، مخاطره ومزالقه التي لابد من التنبه إليها. وهناك عدد من المسائل التي تثيرها العينات منها:

1. **تحديد حجم العينة:**

فلحجم العينة آثاره المباشرة في النتائج المتحصلة، وبصورة عامة كلما كبر هذا الحجم انعكس ذلك بصورة إيجابية على النتائج وأسهم في زيادة الثقة بالعينة والنتائج المتحصلة منها. ولكن كلما كبر حجم العينة تطلب ذلك المزيد من الجهد والوقت والنفقات مما يتعارض مع الفكرة الأساسية الموجهة لأسلوب استخدام العينة. وبطبيعة الحال يجب أن يتناسب حجم العينة مع حجم الأصل أو المجتمع الأصلي الذي ستسحب منه. ولكن لابد من الانتباه إلى أنه ليس ثمة قاعدة ثابتة بصدد عدد أفراد العينة ونسبتهم إلى المجتمع الأصلي مع أن هناك من يرى أن حجم العينة يجب أن يكون ما بين 10 إلى 15٪ من حجم المجتمع الأصلي. ومما يجب تأكيده في هذا السياق أن القرار حول حجم العينة ونسبة تمثيلها للمجتمع الأصلي لا يتحدد فقط بكم المجتمع بل يتحدد أيضاً بطبيعته وصفاته الأساسية، كما يتحدد بأغراض الدراسة ومستوى الدقة المرجوة منها. وهذا يعني أن ثمة عوامل عديدة تؤثر في حجم العينة المختارة، بالإضافة إلى حجم المجتمع الأصلي، منها نسبة الخطأ المسموح به، والأغراض التي يتوخاها الباحث من وراء البحث، والإمكانات الواقعية المتاحة.

2. مسألة انحياز العينة:

إن الباحث هو الذي يقوم بعملية اختيار العينة من المجتمع الأصلي الذي يضمها، وقد يرتكب بعض الأخطاء في أثناء عملية الاختيار، وقد ينحاز بشكل أو بآخر لفئة معينة، ويعطيها وزناً أكبر من الوزن الذي تستحقه. وعموماً لابد أن تتوافر في الباحث صفات الموضوعية والدقة والنزاهة حتى يتجنب الأخطاء أو يقلل من احتمالات ظهورها، ويجب أن يكون أسلوب سحب العينة من المجتمع الأصلي هو أسلوب العشوائية في الاختيار، بحيث تتوافر فرص متكافئة أمام جميع أفراد المجتمع الأصلي للظهور في العينة، وتتضاءل احتمالات التحيز إلى الحدود القصوى.

ومما تجدر الإشارة إليه أن العينة يستحيل أن تكون صورة مصغرة مطابقة للمجتمع الأصلي أو مرآة عاكسة لهذا المجتمع بكافة ملامحه وصفاته مهما توافرت فيها من شروط الموضوعية والدقة والحياد. وهذا يعني أن مسألة تمثيل العينة للمجتمع الأصلي هي مسألة نسبية، وهي تعبير عن درجة ما من درجات التمثيل، وتخضع لمتغيرات عديدة بينها حجم العينة، ونوعها، وأسلوب سحبها، بالإضافة لظروف الحظ وأخطاء القياس. وليس من النادر أن تختلف نتائج دراستين لظاهرة واحدة بسبب اختلاف العينة على الرغم من سلوك الطريق الصحيح في سحبها. إلا أن هذا كله لا يقلل من قيمة العينة وأهميتها، بل يعني أنه لابد من الانتباه إلى كل المتغيرات والعوامل المؤثرة فيها. ومن الطرائق المتبعة للتأكد من تمثيل العينة وصدقها أن يعمد الباحث إلى سحب عدة عينات من المجتمع الأصلي ذاته على أن تتساوى جميعاً في عدد أفرادها، ثم يقوم بمقارنة متوسطات تلك العينات وانحرافاتها المعيارية وغيرها، فإذا دلت تلك المقارنة على عدم وجود فروق ذات دلالة بين تلك العينات تأكد من تمثيل العينة للمجتمع الأصلي الذي اشتقت منه.

طرائق اختيار العينات:

يمكن التمييز بين أربع طرائق في اختيار العينات وهي الطريقة العشوائية، والطريقة الطبقية، والطريقة العرضية، والطريقة المقصودة.

أ. الطريقة العشوائية:

إن السمة الأساسية المميزة للطريقة العشوائية في اختيار العينة هي إتاحة الفرص أمام جميع أفراد المجتمع الأصلي للظهور في العينة وبحيث تكون هذه الفرص متكافئة ولا تفسح مجالاً للانحياز. ومن أبسط الوسائل في اختيار العينة بهذه الطرقة كتابة أسماء جميع أفراد المجتمع على بطاقات صغيرة متشابهة، على ان يتم تقليبها حتى تختلط مع بعضها، ثم نختار العدد المطلوب من هذه الأوراق (البطاقات) دون تمييز. وكثيراً ما تُسمى طريقة الاختيار هذه بطريقة القرعة، كما تُسمى العينة المسحوبة بهذه الطريقة بالعينة العشوائية البسيطة. ومن الواضح أن هذه الطريقة تتطلب جهداً ووقتاً كبيرين. ويمكن استخدام جداول الأعداد العشوائية لتسهيل هذه الطريقة وتسريعها.

ومن مزايا العينة العشوائية المنتظمة أنها تقلل احتمالات الخطأ إلى درجة كبيرة، كما تتسم بسهولة استخدامها وملاءمتها للدراسات التربوية. ويتم الاختيار والسحب في هذا النوع من أنواع العينات بطريقة منتظمة. فبعد تحديد عدد أفراد العينة ونسبتهم إلى المجتمع الأصلي يعمد الباحث إلى إضافة مقدار ثابت هو مقدار "الفاصلة" بطريقة منتظمة إلى الرقم الأول الذي بدأ منه. فإذا كان عدد أفراد المجتمع الأصلي (500) وكان المطلوب هو سحب عينة مؤلفة من (100) فإن مقدار "الفاصلة " هو $\frac{500}{100}=5$ وهو المقدار الثابت الذي يجب أن يُضاف إلى الرقم الأول بصورة منتظمة ومتتابعة مع مراعاة ان يتم اختيار الرقم الأول بالطريقة العشوائية. فإذا كان الرقم الأول الذي تم اختياره كنقطة البداية هو (7) فإن الأرقام التالية التي يجب سحبها هي:

7 12 17 22 27 32 ......الخ

وإذا كان الرقم الأول الذي تم اختياره كنقطة البداية هو 3 فإن الأرقام التالية التي يجب سحبها هي:

3 8 13 18 23 28 وهكذا

ب. الطريقة الطبقية:

وتقوم هذه الطريقة على تقسيم المجتمع الأصلي الواحد إلى عدد من الطبقات، بحيث يسحب من كل طبقة عدد محدد من الأفراد يتلاءم مع نسبة عدد أفراد الطبقة إلى عدد أفراد المجتمع الأصلي. فإذا كان المجتمع الأصلي، على سبيل المثال، يضم ألفاً من الأفراد وكان مؤلفاً من 3 طبقات هي (أ) و(ب) و(ج)، وكانت نسبة كل طبقة إلى المجتمع ككل هي 25٪ و15٪ و60٪ على التوالي. وكان المطلوب هو سحب عينة طبقية مؤلفة من (200) فرداً أي بنسبة 20٪ إلى المجتمع الأصلي ككل. فإنه يتعين على الباحث عند اختياره العينة مراعاة نسبة كل طبقة إلى المجتمع مع المحافظة على النسبة العامة المقررة لسائر الطبقات وهي 20٪. وهذا يعني انه يجب سحب 20٪ من الطبقة الأولى (أ) التي تؤلف 25٪ من المجتمع ككل (250) فرداً أي 50 فردً و20٪ من الطبقة الثانية (ب) التي تؤلف 15٪ من المجتمع ككل (150فرداً) أي 30 فرداً و20٪ من الطبقة الثالثة (ج) التي تؤلف 60٪ من المجتمع ككل 600 فرداً أي 120 فرداً.

ومن الواضح أن العينة العشوائية الطبقية تتطلب التعامل مع كل طبقة وكأنها مجتمع مستقل، وانها تتفوق على العينة العشوائية المتجانسة من حث أن هذه الأخيرة قد تتركز على إحدى الطبقات في المجتمع، وتهمل بقية الطبقات أو تعطيها وزناً أقل من وزنها الحقيقي، وبالتالي فهي لا تعطي صورة صادقة عن الوضع في المجتمع. وكثيراً ما يطلق على الطريقة الطبقية في اختيار العينة اسم الطريقة الطبقية العشوائية أو العينة العشوائية الطبقية نظراً لأن الاختيار ضمن كل طبقة يتم بالطريقة العشوائية ويقوم على توحيد فرص واحتمالات

ظهور أي فرد في العينة. ولهذا النوع من العينات أهميته القصوى في البحوث النفسية والتربوية نظراً لأن هذه البحوث تتصدى، في أغلب الأحيان، للعديد من المتغيرات وتتطلب تصنيف المجتمع إلى طبقات أو فئات استناداً إلى تلك المتغيرات. ومن المتغيرات الهامة في هذا المجال متغير الجنس (الذكورة والأنوثة)، ومتغير العمر، ومتغير الوضع الاجتماعي والاقتصادي، ومتغير الوضع التعليمي أو الحالة التعليمية، ومتغير التحصيل، ومتغير الذكاء، وغيرها.

ج. الطريقة المقصودة:

وتقوم على الاختيار المقصود لعينة من لأفراد اعتماداً على الخبرة السابقة للباحث أو غيره في اختيار العينة من المجتمع الأصلي موضع الاهتمام. فإذا دلت نتائج إحدى الدراسات على أن طلبة أحد أقسام كلية الآداب يمثلون من حيث مستواهم العلمي طلبة هذه الكلية بجميع أقسامها فيمكن الاقتصار على عينة من طلبة هذه القسم واختيارها عن عمد انطلاقاً من أنها تمثل طلبة الكلية ككل. ومن الواضح أنه إذا كان الغرض من الدراسة هو الحصول على تقديرات ونتائج تقريبية فيمكن اللجوء إلى المعاينة المقصودة أو العمدية، أما إذا كان الغرض هو الحصول على نتائج دقيقة وتعميمها، فلابد من اللجوء إلى المعاينة الاحتمالية نظراً لأنها تقلل من احتمالات الخطأ كما تمكننا من قياس الأخطاء والتحكم بها إلى حد بعيد. وبطبيعة الحال فإن الطريقة المقصودة في سحب العينة لا تتعارض مع أسلوب السحب العشوائي، وقد تتطلب هذا الأسلوب إذا كان العدد المطلوب ضئيلاً نسبياً. فإذا كان المطلوب في المثال السابق هو سحب عينة مؤلفة من 100 طالباً من أحد اقسام كلية الآداب فلابد أن يتم السحب في إطار هذا القسم بطريقة المعاينة الاحتمالية العشوائية.

د. الطريقة العرضية:

والفكرة الموجهة لهذه الطريقة هي أنه كثيراً ما تتوافر أمام الباحث عينة مناسبة بمحض المصادفة ودون أن يكون له دخل في إعدادها وتنظيمها. فإذا كان الباحث يقيم على مقربة من إحدى المدارس الثانوية مثلاً، وكان يسعى إلى اختيار عينة من طلاب المرحلة الثانوية فيمكنه أن يقرع أبواب هذه المدرسة التي وفرتها له المصادفة المحضة ويجري الدراسة فيها. ومن فوائد العينة العرضية أنه يمكن عن طريقها تكوين فكرة أولية عن المسألة موضع الدراسة، والحصول على بيانات مفيدة حولها بيسر وسهولة مما يشكل أساساً لصياغة فرضيات لاحقة ومتابعة الدراسة بتعمق ودقة. ولا تصلح العينة العرضية لأن تكون منطلقاً لاستنباط الأحكام وتعميمها، نظراً لأن النتائج لا تتعدى الإطار الضيق الذي خضع له الباحث. وبطبيعة الحال فإن العينة العرضية كالعينة المقصودة، تتطلب استخدام أسلوب السحب العشوائي إذا كان العدد المطلوب لأفرادها أقل من العدد الذي وفرته المصادفة للباحث. فإذا قرع الباحث أبواب المدرسة الموجودة بجوار منزله، كما في المثال السابق، وكان العدد المقرر لأفراد عينته هو 50 تلميذاً من أصل 150 تلميذاً يدرسون في الصف الخامس في هذه المدرسة مثلاً كان لابد له من سحب العدد اللازم عشوائياً من تلاميذ الصف الخامس في هذه المدرسة.

ملاحظات:

لاشك أن الطريقة العشوائية تمثل الطريقة العلمية في اختيار العينة، كما أن العينة العشوائية الطبقية تمثل النموذج أو النوع الأهم من أنواع العينات في البحوث التربوية لأسباب منها: أنها الأقدر على أن تعطي صورة صادقة عن الوضع القائم في المجتمع وتعكسه بطبقاته وفئاته المختلفة. والواقع أن هذه العينة هي الأكثر تحقيقاً لمبدأ السحب العشوائي وما يتطلبه من تأمين فرص متكافئة أمام أفراد المجتمع للظهور في العينة، نظراً لما تفرضه من نظام تصنيفي لمجتمع الدراسة يراعي توزع الطبقات (أو الفئات) العديدة الموجودة فيه وبالنسب

الحقيقية، ومن ثم اختيار النسب المقررة للعينة من كل طبقة بالطريقة العشوائية ودونما انحياز لأي من الأفراد أو الطبقات. فإذا أخذنا بالحسبان أن أغلب المجتمعات موضع الاهتمام في البحوث التربوية هي مجتمعات طبقية أو يمكن إخضاعها للتصنيف أو التقسيم الطبقي (الفئوي)، ظهرت أمامنا أهمية هذا النوع من العينات بصورة واضحة وجلية. ولا يقلل هذا بطبيعة الحال من أهمية سحب عينة متجانسة أو غير طبقية في حالات معينة إذا تعذر تصنيف المجتمع إلى فئات أو طبقات أو رأى الباحث أن دراسته لا تتطلب مثل هذا التصنيف.

إلا أن مبدأ السحب العشوائي وما يتطلبه من تأمين فرص متساوية أمام أفراد المجتمع جميعاً للظهور في العينة قد يتعذر تحقيقه بصورة تامة إذا كان المجتمع الأم الذي ستسحب منه العينة كبيراً للغاية، أو كان أفراده ينتشرون على رقعة جغرافية واسعة، أو كان يتعذر الوصول إليهم لسبب أو لآخر. فالسحب العشوائي للعينة يتطلب، من جملة ما يتطلب، حصر أفراد المجتمع الأصلي جميعهم وترتيبهم بطريقة معينة (كالطريقة الأبجدية والرقمية مثلاً) ليتمكن الباحث فيما بعد من اختيار أفراد عينته عشوائياً بعد تحديد حجم هذه العينة من جانبه. فإذا كان أفراد المجتمع الأصلي يعدّون بعشرات الألوف أو مئات الألوف (كتلاميذ أحد صفوف المرحلة الابتدائية في سوريا مثلاً)، وكانوا يتوزعون على مئات المدارس، وينتشرون في رقعة واسعة جداً، فكيف السبيل إلى اختيار عينة منهم سينتشر أفرادها، إذا روعيت شروط السحب العشوائي بصورة تامة، في رقعة واسعة جداً ويتوزعون في مئات المدارس؟ إن الوصول إلى أفراد هذه العينة سيتطلب جهداً ووقتاً كبيرين للغاية، كما سيتطلب إشراك عدد كبير من العاملين أو المساعدين. ويتنافى هذا الأسلوب بالتأكيد مع مبدأ أساسي تقوم عليه عملية الاختيار أو المعاينة وهو الاقتصاد في الوقت والجهد والنفقات.

والطريقة العملية لاختيار العينة في الحالات التي يكون من العسير فيها اختيار أفراد من صفوف ومدارس متعددة ومنتشرة انتشاراً واسعاً يمكن أن تأخذ الخطوات التالية: (هذا مع الإشارة إلى أن هذه العينة لا تحقق شروط السحب

العشوائي بصورة كافية، ولن تكون بالتالي ممثلة لمجتمع الدراسة، أو سيكون تمثيلها ضعيفاً مما يضع قيوداً أمام الباحث في تعميم النتائج التي سيتوصل إليها):

1. حصر المجتمع الأصلي للدراسة أو المجتمع الأم وفئاته (أو طبقاته) المختلفة وفق المتغيرات التصنيفية مدار البحث (من مثل متغيرات الجنس، أو العمر، أو المنطقة الجغرافية، أو نوع المدرسة (كأن تكون حكومية أو خاصة)، أو المستوى الاجتماعي والاقتصادي... الخ).
2. الاختيار العشوائي لبعض فئات هذا المجتمع كاختيار بعض المناطق الجغرافية بطريقة عشوائية أو مقصودة.
3. الاختيار العشوائي لبعض المدارس من تلك المناطق بالذات (ويراعى في هذا الاختيار تمثيل الأنواع والبيئات المختلفة للمدارس).
4. الاختيار العشوائي لبعض الشعب التي تنتمي إلى الصف الدراسي موضع الاهتمام من تلك المدارس بالذات التي وقع عليها الاختيار كما ذكرنا أعلاه. ويراعى في هذا الاختيار عدد الشعب الموجودة في كل مدرسة، فإذا كان هذا العدد هو 6 شعب في إحدى المدارس و3 شعب في أخرى، وكان العدد المقرر اختياره من الشعب هو $^1/_3$ العدد الكلي للشعب كان لابد من أخذ شعبتين من المدرسة الأولى وواحدة فقط من الثانية عشوائياً، وإذا كان هذا العدد هو شعبة واحدة فقط في إحدى المدارس فيمكن أخذ 3/1 عدد أفراد هذه الشعبة عشوائياً.

ويتضح في ضوء ما سبق أن ما يمكن تسميته بالطريقة العملية في اختيار العينة تتطلب تصنيف المجتمع موضع الاهتمام وفقاً لمتغيرات معينة، كما تتطلب عملية اختزال متتبع لتلك المتغيرات ذاتها بدأت في المثال السابق بالمناطق الجغرافية، ثم تناولت المدارس، فالشعب، فالتلاميذ الأفراد. ولعل هذه الطريقة تلبي بعض شروط السحب العشوائي من خلال هذا الاختزال المتتابع والذي يتم

عشوائياً في مراحله المختلفة بصورة عامة (وقد يعتمد الطريقة المقصودة في إحدى مراحله أو بعضها إذا اقتضت الضرورة).

ونقدم من خلال الجدول التالي مثالاً لعينة طبقية يمكن اختيارها بالطريقة العملية. وتظهر في هذا الجدول النسب المئوية وعدد الأفراد لكل من فئات العينة المختلفة بعد أن تم تصنيفها إلى تلك الفئات في ضوء المتغيرات الثلاثة موضع الدراسة وهي موقع المدرسة (في المدينة أو الريف)، ونوع المدرسة (حكومية أو خاصة)، وجنس التلاميذ (بنين – بنات)، هذا مع الإشارة إلى ان عدد أفراد المجتمع الأصلي للتلاميذ هنا هو 100.000 والعدد المقرر لأفراد العينة هو 1000 أي بنسبة 1٪، وأن نسبة تلاميذ الريف والمدينة هي 60٪ و40٪ على التوالي، ونسبة كل من تلاميذ المدارس الحكومية والخاصة هي 80٪ و20٪ على التوالي، ونسبة الذكور والإناث هي 50٪ لكل منهما.

الجدول رقم (35): النسب المئوية وعدد التلاميذ لعينة مؤلفة من 1000 تلميذ:

الموقع		تلاميذ المدارس الحكومية			تلاميذ المدارس الخاصة			المجموع		المجموع الكلي
		بنين	بنات	المجموع	بنين	بنات	المجموع	بنين	بنات	
المدينة	النسبة	٪16	٪16	٪32	٪4	٪4	٪8	٪20	٪20	٪40
	العدد	160	160	320	40	40	80	200	200	400
الريف	النسبة	٪24	٪24	٪48	٪6	٪6	٪12	٪30	٪30	٪60
	العدد	240	240	480	60	60	120	300	300	600
المجموع	النسبة	٪40	٪40	٪80	٪10	٪10	٪20	٪50	٪50	٪100
	العدد	400	400	800	100	100	200	500	500	1000

ويمكن في ضوء ما سبق القول: إن الطريقة العملية المقترحة تماشي الطريقة العشوائية الطبقية من حيث اعتمادها متغيرات تصنيفية معينة مع

إعطاء الوزن النسبي لكل منها. إلا أنها تخالف هذه الطريقة بدرجة ما حين تعمد إلى الاختزال المتتابع للفئات أو المتغيرات مدار البحث لمواجهة مشكلة الانتشار الواسع والتشتت الهائل لأفراد المجتمع.

الدلالة الإحصائية:

لا يتسع المقام في الكتاب الحالي لدراسة مبادئ الإحصاء الاستنتاجي ومقاييس الدلالة الإحصائية المختلفة بصورة وافية. ولابد لمن يرغب في مثل هذه الدراسة من الرجوع إلى الكتب والمراجع المتخصصة بالإحصاء. وسيكون من المفيد، مع ذلك، التعريف بالدلالة الإحصائية وبعض مقاييسها وبخاصة تلك المقاييس التي تنطوي على أهمية قصوى في البحث التربوي وتستخدم فيه على نطاق واسع.

الخطأ المعياري للعينة:

أشرنا فيما سبق إلى أن المسألة الأهم في نظرية العينات هي مسألة تمثيل العينة للمجتمع الأصلي. وللتحقق من تمثيل العينة للمجتمع الأصلي يمكن من حيث المبدأ مقارنة المقاييس الإحصائية للعينة (من مثل المتوسط والانحراف المعياري) بالمقاييس الإحصائية للمجتمع الأصلي لبيان مدى اقترابها منه أو توافقها معه. إلا أن هذا العمل يتطلب إعادة إجراء الدراسة على المجتمع الأصلي كله لاستخراج المقاييس الإحصائية من هذا المجتمع، وهو أمر مستحيل في أغلب الأحيان ويفوق طاقة الباحث أو الفريق لذي يتولى عملية البحث، كما أنه يفقد الدراسة الأولى التي اعتمدت أسلوب المعاينة معناها إذ تصبح عملاً زائداً بعد أن تحل دراسة الأصل محل دراسة الفرع. والواقع أنه نظراً لعدم توافر المقاييس البارامترية (أي المقاييس الفعلية) للمجتمع فإن علماء الإحصاء يعمدون إلى تقديرها من المقاييس المقابلة لها من العينات، ويلجؤون من أجل ذلك إلى سحب عدة عينات متساوية في الحجم من المجتمع الأصلي الواحد، ثم استخراج المقاييس الإحصائية لهذه العينات وحساب الفروق بينها، فإذا لم تظهر بينها فروق دالة

إحصائياً فهذا يعني أنها متجانسة وتنتمي إلى مجتمع أو أصل واحد، وبالتالي فإن كلاً منها هي عينة ممثلة له. فكيف يتم استخراج دلالة الفروق بين المقاييس الإحصائية المختلفة؟ إن هذا العمل يتطلب أولاً حساب الخطأ المعياري لتلك المقاييس.

1) الخطأ المعياري للمتوسط:

لو افترضنا أن لدينا (50) عينة متساوية من حيث العدد كما سحبت من أصل واحد، وإننا قمنا بحساب المتوسط لكل منها، فإننا سنحصل في هذه الحالة على 50 متوسطاً. وهذه المتوسطات الخمسون التي حصلنا عليها لا تتساوى عادة في قيمها ويفترض أن تتوزع قيمها اعتدالياً أي تأخذ شكل المنحنى الاعتدالي نظراً لأن العينات التي استخرجت منها هذه المتوسطات اشتقت من المجتمع يفترض أن يتوزع أفراده بصورة اعتدالية في السمة موضع القياس.

إن المتوسطات الخمسين التي تم استخراجها يمكن استخراج متوسطها (وهو متوسط المجتمع ككل)، كما يمكن استخراج انحرافها المعياري الذي سيشير في هذه الحالة إلى متوسط انحرافات هذه المتوسطات عن متوسطها. وهذا الانحراف المعياري للمتوسطات هو ما يسمى عادة "الخطأ المعياري" بهدف تمييزه عن الانحراف المعياري للدرجات. وعلى ذلك يمكن القول: إن الخطأ المعياري للمتوسطات (والمقاييس الإحصائية الأخرى أيضاً) "يدل على مدى الخطأ المحتمل لتلك المقاييس في ابتعادها أو اقترابها من أصلها الذي انتزعت منه" (البهي للسيد، 1978، ص309).

ويُحسب الخطأ المعياري للمتوسط بقسمة الانحراف المعياري للعينة على الجذر التربيعي لعدد أفرادها وذلك وفق الصيغة الإحصائية التالية:

$$ع م = \frac{ع}{\sqrt{ن}}$$

حيث يشير الرمز ع م إلى الخطأ المعياري للمتوسط.

والرمز ع إلى الانحراف المعياري للعينة.

والرمز ن إلى عدد أفراد العينة.

ويمكن استناداً إلى قيمة الخطأ المعياري للمتوسط تحديد المدى الذي يقع ضمنه المتوسط الحقيقي للمجتمع الأصلي. ويتم ذلك بضرب قيمة الخطأ المعياري للمتوسط بـ 1.96 لمستوى من الثقة قدره 95٪ وبـ2.58 لمستوى من الثقة قدره 99٪ ثم إضافة القيمة الناتجة من عملية الضرب السابقة إلى المتوسط وطرحها منه. وما يسوّغ القيام بالعمل السابق هو أن 95٪ من الحالات في المنحني الاعتدالي تقع ما بين الدرجة المعيارية (الذالية) ± 1.96، في حين ان 99٪ من الحالات في هذا المنحني تقع ما بين الدرجة المعيارية (الذالية)± 2.58. فإذا قبل الباحث بدرجة من الثقة قدرها 95٪ (أي بنسبة خطأ 5٪) تعين عليه ضرب قيمة الخطأ المعياري المحسوب بـ ±1.96، وإذا اختار درجة من الثقة قدرها 99٪ (أي بنسبة خطأ قدرها 1٪ من الحالات) توجب عليه ضرب قيمة الخطأ المشار إليها بـ ± 2.58. وفي المثال السابق نحتاج إلى اتباع الخطوات التالية لتقدير المدى الذي يقع ضمنه المتوسط الحقيقي:

أولاً: ضرب قيمة الخطأ المعياري للمتوسط وهي 0.932 بـ 1.96 لدرجة من الثقة قدرها 95٪ والنتيجة المتحصلة هنا هي 1.809 وبـ 2.58 لدرجة من الثقة قدرها 99٪ والنتيجة المتحصلة هي 2.381.

ثانياً: إضافة هذه القيمة إلى المتوسط وقدره 55 ثم طرحها منه أي 55± 1.809 لمستوى الثقة 95٪ و55 ± 2.381 لمستوى الثقة 99٪. ويمكن بناء على هذه النتيجة القول: إن المتوسط الحقيقي للمجتمع الأصلي يتراوح ما بين 53.191 و56.809 بدرجة من الثقة قدرها 95٪. كما يمكن القول بدرجة من الثقة قدرها 99٪ إنه يتراوح ما بين 52.691 و57.381.

2) الخطأ المعياري للوسيط:

لا تختلف الطريقة المعتمدة في حساب الخطأ المعياري للوسيط من حيث جوهرها عن الطريقة الخاصة بحساب الخطأ المعياري للمتوسط، نظراً لأن الوسيط يتطابق مع المتوسط في التوزيع الاعتدالي. وتتضمن المعادلة الخاصة بحساب الخطأ المعياري للوسيط تعديلاً طفيفاً لمعادلة الخطأ المعياري للمتوسط وتأخذ هذه المعادلة الشكل التالي:

$$ع_و = 1.253 \times \frac{ع}{\sqrt{ن}}$$

حيث يدل الرمز (ع) وعلى الخطأ المعياري للوسيط

و (ع) على الانحراف المعياري للعينة.

و (ن) على عدد أفراد العينة.

فإذا كان وسيط درجات مادة العلوم لدى عينة من التلاميذ عددهم 200 هو 48 والانحراف المعياري هو 12 فإن قيمة الخطأ المعياري للوسيط في هذه الحالة هي:

$$1.063 = \frac{12}{14.142} \times 1.253 = \frac{12}{\sqrt{200}} \times 1.253$$

وحدود هذا الوسيط عند مستوى الثقة 95٪ هي:

الوسيط + الخطأ المعياري ×1.96

الوسيط – الخطأ المعياري × 1.96

أي أن قيمة الوسيط في مثالنا تتراوح ما بين 48 + (1.96× 1.063) و48 – (1.96×1.063) وبذلك يقع الوسيط في المدى من 45.917 إلى 50.083. وأما حدود هذا الوسيط عند مستوى الثقة 99٪ فهي:

الوسيط + الخطأ المعياري × 2.58

الوسيط – الخطأ المعياري × 2.58

وبالتعويض عن قيم المثال السابق يقع الوسيط ضمن المدى من 45.257 إلى 50.742

3) الخطأ المعياري للانحراف المعياري:

والمعادلة الخاصة بحساب الخطأ المعياري للانحراف المعياري هي:

$$ع ع = \frac{ع}{\sqrt{2 ن}}$$

حيث يدل الرمز ع ع على الخطأ المعياري للانحراف المعياري.

والرمز ع على الانحراف المعياري.

والرمز ن على عدد أفراد العينة.

وفي مثالنا السابق حيث عدد أفراد العينة هو 200 والانحراف المعياري هو 12 يكون الخطأ المعياري للانحراف المعياري هو:

$$0.6 = \frac{12}{\sqrt{200 \times 2}}$$

وحدود هذا الانحراف المعياري عند مستوى الثقة 95٪ هي:

الانحراف المعياري + الخطأ المعياري ×1.96

الانحراف المعياري – الخطأ المعياري ×1.96

$12 + (1.96 \times 0.6) = 13.176$

$12 - (1.96 \times 0.6) = 10.824$

أي يقع الانحراف المعياري في المدى من 10.824 إلى 13.176.

وأما عند مستوى الثقة 99٪ فإن حدود الانحراف المعياري في مثالنا هي:

$$13.548=(2.58\times0.6)+12$$

$$10.452=(2.58\times0.6)-12$$

اي يقع الانحراف المعياري في المدى من 10.452 إلى 13.548 عند مستوى الثقة 99٪.

4) الخطأ المعياري للنسبة:

ويرتكز إلى المبدأ ذاته الذي ترتكز إليه الأخطاء المعيارية للمتوسط والوسيط والانحراف المعياري، ويشير إلى مدى الخطأ المحتمل للنسبة. والطريقة المتبعة لحساب الخطأ المعياري للنسبة تقوم على ضرب النسبة بباقي النسبة من الواحد الصحيح وقسمة الناتج على عدد أفراد العينة، ثم حساب الجذر التربيعي للناتج من هذه القسمة، وذلك كما في الصيغة التالية:

$$ع_{أ} = \sqrt{\frac{أ \times ب}{ن}}$$

حيث يشير الرمز ع أ إلى الخطأ المعياري للنسبة أ

ويشير الرمز أ إلى النسبة المراد تقدير خطئها المعياري.

ويشير الرمز ب إلى باقي النسبة من الواحد الصحيح اي 1 – أ.

ويشير الرمز ن إلى عدد أفراد العينة.

وعندما تكون النتائج على شكل نسب مئوية تأخذ المعادلة السابقة الشكل التالي:

$$\text{الخطأ المعياري للنسبة المئوية} = 100 \sqrt{\frac{\text{النسبة المئوية} \times \text{الباقي من مئة}}{\text{عدد أفراد العينة}}}$$

فإذا كانت نسب الإجابات الصحيحة على أحد البنود (الأسئلة) في اختبار ما هي 0.6 وكان عدد الأفراد المفحوصين هو 200 فإن الخطأ المعياري لهذه النسبة هو:

$$0.034 = \sqrt{\frac{(0.6 - 1) \times 0.6}{200}}$$

وأما الخطأ المعياري للنسبة المئوية في هذا المثال فهو:

$$3.464 = 0.034 \times 100$$

وفيما يتصل بحدود النسبة عند مستوى الدلالة 95٪ فإن هذه الحدود هي:

$$0.666 = (1.96 \times 0.034) + 0.6$$

$$0.533 = (1.96 \times 0.034) - 0.6$$

وأما حدود النسبة عند مستوى الدلالة 99٪ فهي:

$$0.687 = (2.85 \times 0.034) + 0.6$$

$$0.512 = (2.58 \times 0.034) - 0.6$$

وبطبيعة الحال ستأخذ النتائج السابقة صورة النسب المئوية في حال استخدام النسبة المئوية. وسوف يقع مدى النسبة المئوية في المثال السابق ما بين

53.3 و66.6 عند مستوى الثقة 95٪ كما سيقع ما بين 51.2 و68.7 عند مستوى الثقة 99٪

5) الخطأ المعياري لمعامل الارتباط:

يحسب الخطأ المعياري لمعامل الارتباط بالمعادلة التالية:

$$ع ر = \frac{1-ر^2}{\sqrt{ن-1}}$$

حيث يشير الرمز ع ر إلى الخطأ المعياري لمعامل الارتباط

والرمز ر إلى معامل الارتباط

والرمز ن إلى عدد أفراد العينة.

فإذا بلغ الارتباط بين اختبار في العلوم واختبار في الرياضيات طبقا على عينة مؤلفة من 100 تلميذ 0.55 فإن الخطأ المعياري لهذا المعامل هو:

$$\frac{1-(0.55)^2}{\sqrt{100-1}} = \frac{0.697}{9.949} = 0.07$$

وتقع قيمة معامل الارتباط في هذا المثال بين 0.55 + (0.07 × 1.96) أي 0.687 و0.55 – (0.07 × 1.96) أي 0.413 عند مستوى الدلالة 95٪ (أي يتراوح من 0.413 إلى 0.687).

وأما قيمة هذا المعامل عند مستوى الدلالة 99٪ فتقع بين 0.55 + (0.07 × 2.58) أي 0.73 و0.55 – (0.07 × 2.58) أي 0.37 (أي تتراوح بين 0.37 و0.73).

بعض مقاييس الدلالة الإحصائية:

ما من شك في أن مقاييس الدلالة الإحصائية تنطوي على أهمية فائقة للباحث لما توفره له من إمكان التأكد من صحة نتائجه وما إذا كانت الفروق التي أظهرتها هذه النتائج فروقاً حقيقية أم أنها عائدة إلى عامل المصادفة والحظ.

وسنتوقف فيما يلي عند اثنين فقط من أهم مقاييس الدلالة الإحصائية وأكثرها شيوعاً واستخداماً، وهما اختبار (ت) ومقياس (كاي2) لحسن المطابقة، كما سنقدم وصفاً سريعاً لدلالة الفرق بين الانحرافات المعيارية، ودلالة الفرق بين النسب، إضافة إلى الدلالة الإحصائية لمعامل الارتباط، تاركين لمن يرغب في الدراسة الموسعة للدلالة الإحصائية ومقاييسها فرصة العودة إلى المراجع الإحصائية المتخصصة.

اختبار (ت) لدلالة الفروق بين المتوسطات:

يطلق على الاختبار الخاص بالكشف عن الفروق بين المتوسطات اسم اختبار (ت) للإشارة إلى أكثر الحروف تكراراً في اسم واضعه ستودنت وهو حرف التاء. والهدف من هذا الاختبار: هو التأكد مما إذا كانت الفروق الظاهرة بين المتوسطات فروقاً حقيقية وثابتة، أم أنها ناجمة عن الصدفة وظروف اختيار العينة. ولهذا الاختبار استعمالاته الواسعة في البحوث النفسية والتربوية، فهو يستخدم لقياس دلالة فروق المتوسطات المرتبطة وغير المرتبطة سواء للعينات المتساوية أم للعينات غير المتساوية. والصيغة الخاصة بحساب (ت) في حالة استخدام عينتين متساويتي العدد وغير مرتبطتين هي التالية:

$$\text{ت} = \frac{\text{م}_1 - \text{م}_2}{\sqrt{\frac{\text{ع}_1^2 + \text{ع}_2^2}{\text{ن} - 1}}}$$

حيث يشير الرمز م1 إلى متوسط المجموعة الأولى

والرمز م2 إلى متوسط المجموعة الثانية

والرمز ع1 إلى الانحراف المعياري للمجموعة الأولى

والرمز ع2 إلى الانحراف المعياري للمجموعة الثانية

والرمز ن إلى عدد أفراد العينة الأولى أو الثانية (فهما متساويتان كما أشرنا).

فلو كان متوسط الدرجات للمجموعة الأولى هو 12 ومتوسط الدرجات للمجموعة الثانية هو 10.

والانحراف المعياري للمجموعة الأولى هو 4، والانحراف المعياري للمجموعة الثانية هو 5 وعدد أفراد كل من هاتين المجموعتين هو 10.

فهل هناك فرق ذو دلالة إحصائية بين متوسطي المجموعتين؟ للإجابة على هذا السؤال لابد من حساب قيمة (ت) على النحو التالي:

$$ت = \frac{12 - 10}{\sqrt{\frac{(4)^2 + (5)^2}{10 - 1}}}$$

ولبيان ما إذا كانت قيمة (ت) المستخرجة في مثالنا دالة إحصائياً أم لا علينا أولاً تحديد درجات الحرية، ودرجات الحرية هي عدد الدرجات أو الفئات — عدد القيود الإحصائية وهي في مثالنا 10 – 1=9، وعلينا ثانياً النظر في جدول دلالة (ت) عند درجة الحرية 9 ونسبة الخطأ 0.05 و0.01 و0.001 فإذا كانت قيمة (ت) في الجدول أقل من قيمة (ت) المستخرجة في مثالنا عند أي من النسب الثلاث السابقة كان الفرق دالاً عند تلك النسب.

الجدول (36): جدول دلالة (ت):

د.ج	0.05	0.01	0.001	د.ج	0.05	0.01	0.001
1	12.706	63.657	639.619	18	2.101	2.878	3.922
2	4.352	9.925	30.598	19	2.093	2.861	3.883
3	3.182	5.841	22.941	20	2.086	2.845	3.850
4	2.776	4.604	8.610	21	2.080	2.830	3.819
5	2.571	4.032	6.859	22	2.074	2.819	3.792
6	2.447	3.770	5.459	23	2.069	2.807	3.767
7	2.365	3.499	5.405	24	2.064	2.797	3.745
8	2.306	3.355	5.041	25	2.060	2.787	3.725
9	2.262	3.250	4.780	26	2.056	2.779	3.707
10	2.228	3.169	4.587	27	2.052	2.771	3.690
11	2.201	3.106	4.137	28	2.048	2.763	3.674
12	2.189	3.055	4.318	29	2.045	2.756	3.659
13	2.160	3.012	4.321	30	2.032	2.750	3.646
14	2.145	2.977	4.140	40	2.02	2.704	3.551
15	2.131	2.947	4.073	60	2.00	2.660	3.46
16	2.120	2.921	4.015	120	1.980	2.617	3.373
17	2.110	2.898	3.965	فما فوق	1.960	2.576	3.291

وبالنظر إلى الجدول السابق نجد أن قيمة (ت) عند درجة الحرية 9 هي 2.262 و3.250 و4.780 لكل من النسب الثلاث 0.05 و0.01 و0.001 على التوالي، وهذه القيم أعلى من قيمة (ت) المحسوبة في مثالنا مما يدل على أنه لا يوجد فرق ذو دلالة بين المتوسطين.

ومن الصيغ الأخرى الهامة لاختبار (ت) الصيغة الخاصة بحساب (ت) في حال استخدام عينتين غير متساويتي العدد وغير مرتبطتين، وتظهر هذه الصيغة فيما يلي:

$$\text{ت} = \frac{\text{م}_1 - \text{م}_2}{\sqrt{\left(\frac{\text{ن}_1 \text{ع}_1^2 + \text{ن}_2 \text{ع}_2^2}{\text{ن}_1 + \text{ن}_2 - 2}\right)\left(\frac{1}{\text{ن}_1} + \frac{1}{\text{ن}_2}\right)}}$$

حيث يشير الرمز م_1 إلى متوسط المجموعة الأولى

والرمز م_2 إلى متوسط المجموعة الثانية

والرمز ع_1 إلى الانحراف المعياري للمجموعة الأولى

والرمز ع_2 إلى الانحراف المعياري للمجموعة الثانية

والرمز ن_1 إلى عدد أفراد المجموعة الأولى

والرمز ن_2 إلى عدد أفراد المجموعة الثانية

وبالإضافة للصيغتين السابقتين هناك صيغ أخرى لحساب دلالة الفروق بين المتوسطات باختبار (ت). إلا أن هذه الصيغ جميعها ترتكز إلى مبدأ أو أساس واحد وهو نسبة الفرق بين المتوسطات إلى الخطأ المعياري لهذا الفرق.

دلالة الفروق بين الانحرافات المعيارية:

قد يحتاج الباحث في التربية في حالات معينة إلى الكشف عن دلالة الفروق بين مجموعتين أو (أكثر) من حيث تشتت (أو تباين) الدرجات لكل مجموعة. إنه في مثل هذه الحالات يسعى إلى الكشف عما إذا كانت الفروق في الانحرافات المعيارية دالة إحصائياً بدلاً من اهتمامه بالفروق بين المتوسطات

والدلالة الإحصائية لتلك الفروق. وتشبه الطريقة العامة لحساب دلالة الفروق بين الانحرافات المعيارية الطريقة العامة لحساب دلالة الفرق بين المتوسطات، وتقوم على قسمة الفرق بين الانحرافات المعيارية على الخطأ المعياري لهذا الفرق. ومن الصيغ الإحصائية المستخدمة لحساب دلالة الفروق بين الانحرافات المعيارية الصيغة التالية التي تصلح بخاصة للعينات غير المرتبطة وكبيرة العدد.

$$\text{دلالة الفروق بين الانحرافات المعيارية} = \frac{ع_1 - ع_2}{\sqrt{عع_1^2 + عع_2^2}}$$

حيث يشير الرمز ع إلى الانحراف المعياري.

والرمز $عع_1$ إلى الخطأ المعياري للانحراف المعياري للمجموعة الأولى

والرمز $عع_2$ إلى الخطأ المعياري للانحراف المعياري للمجموعة الثانية

والخطوات اللازمة لتطبيق الصيغة السابقة يمكن إيضاحها من خلال المثال التالي:

لو طبقنا اختباراً في العلوم على مجموعة مؤلفة من 150 طالبة ومجموعة أخرى مؤلفة من 100 طالب وكان الانحراف المعياري لدرجات الإناث 12 ولدرجات الذكور 8. فهل الفرق في الانحراف المعياري بين المجموعتين دال إحصائياً؟

للإجابة على هذا السؤال لابد من اتباع الخطوات التالية:

1. حساب الخطأ المعياري للانحراف المعياري للمجموعة الأولى (مجموعة الإناث) باستخدام المعادلة الخاصة بحساب الخطأ المعياري للانحراف المعياري وهي:

$$ع ع = \frac{ع}{\sqrt{2 ن}}$$

وهذا الخطأ في مثالنا هو:

$$\frac{12}{\sqrt{150 \times 2}} = 0.692$$

2. حساب الخطأ المعياري للانحراف المعياري للمجموعة الثانية (مجموعة الذكور) باستخدام المعادلة السابقة نفسها. وهذا الخطأ في مثالنا هو:

$$\frac{8}{\sqrt{100 \times 2}} = 0.565$$

3. تطبيق المعادلة الخاصة بحساب دلالة الفروق بين الانحرافات المعيارية. وبالتعويض عنها في مثالنا نحصل على القيم التالية:

$$0.479 = \frac{8 - 12}{\sqrt{(0.565)^2 + (0.692)^2}}$$

وهذه القيمة أعلى من 1.96 وهو مستوى الدلالة عند 0.05 كما أنها أعلى من 2.58 وهو مستوى الدلالة عند 0.01 مما يدل على أن الفرق في الانحراف المعياري بين المجموعتين دال إحصائياً عند مستويي الدلالة 0.05 و0.01 ويؤكد بالتالي أن الفرق في التشتت بين درجات البنات ودرجات البنين هو فرق حقيقي.

إلا أن المعادلة السابقة تصلح للعينات الكبيرة كما أسلفنا. ولحساب دلالة الفرق في الانحراف المعياري بين العينات الصغيرة (أي أقل من 30 فرداً) يمكن استخدام اختبار (ف) لفيشر أو ما يعرف بالنسبة الفائية أو التباينية. ويتم ذلك بقسمة التباين الكبير (والتباين هو مربع الانحراف المعياري) على التباين الصغير، أي:

$$\text{النسبة الفائية} = \frac{\text{التباين الكبير}}{\text{التباين الصغير}}$$

ويتضح ذلك في المثال التالي:

إذا كان عدد أفراد المجموعة الأولى 10

وعدد أفراد المجموعة الثانية 15

والتباين في المجموعة الأولى 18

والتباين في المجموعة الثانية 30

فإن النسبة التباينية هي:

$$\frac{30}{18} = 1.666$$

وبالرجوع إلى الجداول الإحصائية الخاصة بدلالة (ف) عدد درجات الحرية 15 – 1=14 للتباين الكبير (المجموعة الثانية) و10 – 1=9 للتباين الصغير (المجموعة الأولى). نجد ان هذه النتيجة غير دالة إحصائياً عند مستويي الدلالة 0.05 و0.01 مما يسمح بالقول: إن الفرق في التباين بين المجموعتين غير حقيقي.

دلالة الفروق بين النسب:

ويقوم استخراج دلالة الفروق بين النسب على المبدأ ذاته الذي يقوم عليه استخراج الفروق للمقاييس الإحصائية المختلفة، ويهدف بدوره إلى الكشف عما إذا كانت الفروق بين النسب دالة إحصائياً عند مستوى معين للثقة.

والطريقة الخاصة بحساب دلالة الفروق بين النسب والتي تصلح بخاصة للعينات غير المرتبطة وغير المتساوية، تقوم على حساب الفرق بين النسبتين، ثم قسمة الناتج على الخطأ المعياري لهذا الفرق. ويمكن إيضاح هذه الطريقة بالمثال التالي:

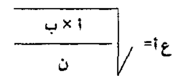
طبق اختبار في اللغة الإنكليزية على مجموعة من الذكور عدد أفرادها 100 ومجموعة من الإناث عدد أفرادها 80 وكانت نسبة من أعطى إجابات صحيحة على أحد أسئلة هذا الاختبار هي 60٪ من الذكور و40٪ من الإناث فما دلالة هذا الفرق بين النسبتين؟

ويمكن اتباع الخطوات التالية للإجابة على السؤال السابق:

1) حساب الفرق بين النسبتين وهو في هذا المثال:

(0.20) ٪20 = ٪40 – ٪60

2) حساب الخطأ المعياري لكل من النسبتين على حدة وذلك وفق الصيغة التالية (والتي شرحت سابقاً):



$$ع_أ = \sqrt{\frac{أ \times ب}{ن}}$$

حيث:

أ: هي النسبة المراد تقدير خطئها المعياري.

و ب هي باقي النسبة من الواحد الصحيح أي 1 – أ

و ن عدد أفراد العينة.

وبتطبيق هذه الصيغة على عينة الذكور نحصل على الخطأ المعياري لعينة الذكور وهو:

$$\sqrt{\frac{0.4 \times 0.6}{100}} = 0.048$$ أي 4.8%

وبالطريقة نفسها نحصل على الخطأ المعياري لعينة الإناث وهو:

$$\sqrt{\frac{0.6 \times 0.4}{80}} = 0.054 \text{ اي } 5.4\%$$

3) حساب الخطأ المعياري للفرق بين النسبتين وهو الجذر التربيعي لحاصل جمع الخطأ المعياري للنسبة الأولى والخطأ المعياري للنسبة الثانية. والصيغة الخاصة بحسابه هي:

$$\sqrt{\text{ع ن}_1^2 + \text{ع ن}_2^2}$$

ويتطبيق هذه الصيغة على المثال السابق يكون الخطأ المعياري للفرق بين النسبتين هو:

$$\sqrt{(0.048)^2 + (0.054)^2} = 0.072$$

وبالانتهاء من الخطوات السابقة، لابد من قسمة الفرق بين النسبتين وهو 0.20 على الخطأ المعياري للفرق وهو 0.072 وناتج هذه القسمة هو 2.777.

ويمكن معرفة دلالة هذا الفرق عند مستويات مختلفة من الثقة بالرجوع إلى الجداول الإحصائية الخاصة بدلالة فروق النسب. وبالإضافة للصيغة السابقة هناك صيغ أخرى لحساب دلالة الفروق بين النسب للعينات المرتبطة والصغيرة.

الدلالة الإحصائية لمعامل الارتباط:

الهدف من وراء حساب الدلالة الإحصائية لمعامل الارتباط هو الكشف عما إذا كان الارتباط القائم بين المتغيرين موضع الدراسة ارتباطاً حقيقياً أم لا. ويرتكز حساب الدلالة الإحصائية لمعامل الارتباط على فكرة الخطأ المعياري. وبما أن الخطأ المعياري للارتباط " يدلّ على الانحراف المعياري لتوزيع معاملات

الارتباط فإن الخطأ المعياري الذي يمتد إلى 1.96 درجة معيارية = 1.96 × 0.1 = 0.196 " (البهي السيد، 1978، ص326). وذلك باعتبار أن التوزيع الاعتدالي يتراوح من – 5 إلى +5 درجة معيارية (أي يتضمن 10 درجات معيارية). وتبعاً لذلك فإنه إذا كانت القيمة العددية لمعامل الارتباط المحسوب أكبر من 0.196 فإن بوسعنا أن نقرر أن هذا الارتباط هو ارتباط حقيقي بنسبة 95٪ ثقة ونسبة 5٪ شك).

وللسبب نفسه فإنه إذا كانت القيمة العددية لمعامل الارتباط المحسوب أكبر من 0.258 فإن بوسعنا أن نقرر أن هذا الارتباط هو ارتباط حقيقي (بنسبة 99٪ ثقة و1٪ شك).

وقد أعدت جداول بالدلالة الإحصائية للارتباط الذي تزيد قيمته على الصفر استندت إلى جداول مساحات المنحنى الاعتدالي المعياري. وبفضل تلك الجداول لا يتطلب حساب الدلالة الإحصائية لمعامل الارتباط سوى معرفة عدد أفراد العينة، ثم حساب درجات الحرية، وهي في هذه الحالة عدد أفراد العينة – 2 (والرقم 2 يشير إلى عدد القيود الإحصائية أي المتغيرين اللذين يدرس الارتباط بينهما)، ثم النظر في الجداول الخاصة بدلالة معاملات الارتباط أمام درجة الحرية المحسوبة وتحت النسبتين 0.05 و0.01. ومن الواضح أنه إذا كان معامل الارتباط المحسوب مساوياً أو أكبر من القيمة الموجودة تحت نسبة 0.01 قلنا عنه: إنه دال إحصائياً عند هذه النسبة أي أن نسبة الشك فيه هي 1٪ فقط. أما إذا كان معامل الارتباط المحسوب مساوياً أو أكبر من القيمة الموجودة تحت نسبة 0.05 فقط قلنا عنه: إنه دال عند هذه النسبة فقط (أي أن نسبة الشك فيه هي 5٪). وبالطبع فإنه إذا كان معامل الارتباط المحسوب أقل من القيمة الموجودة تحت أي من هاتين النسبتين قلنا عنه: إنه غير دال إحصائياً.

فإذا كان عدد أفراد العينة 82

وكان معامل الارتباط 0.26

فإن درجات الحرية = 82 – 2= 80

وبالنظر في جداول دلالة معامل الارتباط عند درجة الحرية 80 وبنسبة 0.05 و0.01 شكاً يتبين ان القيمة المحسوبة أعلى من القيمة الموجودة تحت 0.05 واقل من القيمة الموجودة تحت 0.01. واستناداً إلى ذلك يمكن القول: إن معامل الارتباط دل إحصائياً عند مستوى 0.05 فقط، وبالتالي فإن الارتباط حقيقي بنسبة 95٪ ثقة و5٪ شكاً.

الجدول (37): جدول دلالة معامل الارتباط:

درجات الحرية ن – 2	95٪ ثقة 5٪ شك	99٪ ثقة 1٪ شك	درجات الحرية ن – 2	95٪ ثقة 5٪ شك	99٪ ثقة 1٪ شك
1	0.997	1.000	24	0.388	0.496
2	0.950	0.990	25	0.381	0.487
3	0.878	0.951	26	0.374	0.478
4	0.811	0.917	27	0.367	0.470
5	0.754	0.874	28	0.361	0.463
6	0.707	0.834	29	0.355	0.456
7	0.666	0.798	30	0.349	0.449
8	0.632	0.765	35	0.325	0.418
9	0.602	0.735	40	0.304	0.393
10	0.576	0.708	45	0.288	0.372
11	0.553	0.684	50	0.273	0.354
12	0.532	0.661	60	0.250	0.325
13	0.514	0.641	70	0.233	0.302
14	0.497	0.623	80	0.217	0.283
15	0.482	0.606	90	0.205	0.267
16	0.468	0.590	100	0.195	0.254
17	0.456	0.575	125	0.174	0.228

درجات الحرية ن – 2	95% ثقة 5% شك	99% ثقة 1% شك	درجات الحرية ن – 2	95% ثقة 5% شك	99% ثقة 1% شك
18	0.444	0.561	150	0.159	0.208
19	0.433	0.549	200	0.138	0.181
20	0.423	0.537	300	0.113	0.148
21	0.413	0.526	400	0.098	0.128
22	0.404	0.515	500	0.088	0.115
23	0.396	0.505	1000	0.062	0.081

إلا أنه لا يصح اعتماد الخطأ المعياري في الكشف عن الدلالة الإحصائية لمعامل الارتباط حين تكون قيمة الارتباط مرتفعة، ذلك أن ارتفاع هذه القيمة يؤدي إلى التواء توزيع معاملات الارتباط، وينطبق هذا أيضاً على الدلالة الإحصائية للفرق بين معاملات الارتباط. ولمواجهة هذه المشكلة عمدّ فيشر إلى إعداد جداول خاصة للمقابلات اللوغاريتمية ل (ر)

ولحساب دلالة الفرق بين معاملات الارتباط تستخدم المعادلة التالية:

$$\text{معادلة دلالة الفرق بين معاملات الارتباط} = \frac{ز_1 - ز_2}{\sqrt{\frac{1}{ن_1 - 3} + \frac{1}{ن_2 - 3}}}$$

حيث يدل الرمز $ز_1$ على المقابل اللوغاريتمي لمعامل الارتباط في المجموعة الأولى.

ويدل الرمز $ز_2$ على المقابل اللوغاريتمي لمعامل الارتباط في المجموعة الثانية.

و$ن_1$ إلى عدد أفراد المجموعة الأولى.

و $ن_2$ إلى عدد أفراد المجموعة الثانية.

فإذا قام الباحث بتطبيق اختبارين أحدهما في الرياضيات، والثاني في العلوم على مجموعة من الطلاب عددهم 80 طالباً، ومجموعة أخرى من الطالبات عددهن 100 طالبة، وقام بحساب معامل الارتباط بين الاختبارين في كل مجموعة على حدة، وبلغ هذا الارتباط في مجموعة الطلاب 0.64، وفي مجموعة الطالبات 0.48، فهل الفرق بين معاملي الارتباط في المجموعتين دال إحصائياً؟

للإجابة على هذا السؤال لابد من اتباع الخطوتين التاليتين:

1. الكشف عن المقابل اللوغاريتمي لمعاملي الارتباط في الجداول الخاصة بذلك وهي في هذا المثال 0.76 لمعامل الارتباط 0.64 و0.52 لمعامل الارتباط 0.48.
2. تطبيق المعادلة السابقة.

وبالتعويض عن المعادلة السابقة نحصل على:

$$\frac{0.52-0.76}{\sqrt{\frac{1}{3-100}+\frac{1}{3-80}}} = \frac{0.24}{0.152} = 1.57 \text{ تقريباً}$$

إن القيمة التي حصلنا عليها أقل من القيمة الواقعة عند نسبة الثقة 95٪ و99٪ مما يشير إلى أن الفرق بين معاملي الارتباط في مجموعتي الطلاب والطالبات غير دال إحصائياً.

مقياس كاي^2 لحسن المطابقة:

مقياس كاي^2 هو من أهم مقاييس الدلالة الإحصائية، وقد أطلق عليه هذا الاسم نسبة إلى واضعه كارل بيرسون. والغرض من هذا المقياس: هو الكشف عن مدى اختلاف التكرار التجريبي أو الواقعي عن التكرار المتوقع أو المحتمل،

وبالتالي التأكد مما إذا كانت القيم أو التكرارات التجريبية تختلف اختلافاً دالاً عن القيم أو التكرارات المتوقعة التي يفترض حدوثها على المستوى الاحتمالي النظري. وتزداد قيمة كاي^2 مع ازدياد الفرق بين التكرار التجريبي والتكرار المتوقع أو المحتمل وتنخفض هذه القيمة مع تناقص هذا الفرق إلى أن تنعدم (أي تصل إلى الصفر) بانعدامه مما يشير في هذه الحالة الأخيرة إلى مطابقة التكرار التجريبي للتكرار المتوقع (وهو التكرار الاعتدالي)، ويؤكد أن التوزيع أصبح اعتدالياً وترتكز الطريقة العامة لحساب كاي^2 على الصيغة التالية:

$$كا^2 = مج \frac{(ت و - ت م)^2}{ت م}$$

حيث يشير الرمز ت و إلى التكرار الواقعي (أو التجريبي أو الملاحظ)

والرمز ت م إلى التكرار المتوقع (الاحتمالي)

فإذا رمينا قطعة معدنية من النقود في الهواء خمسين مرة، فإننا نتوقع أن تظهر الكتابة 25 مرة والصورة 25 مرة. إلا أن هذا التوقع قد يكون مخالفاً للواقع، وقد تظهر الكتابة في الواقع 20 مرة والصورة 30 مرة وفي هذه الحالة فإن الفرق بين ما هو متوقع وما حدث فعلاً فيما يتصل بالكتابة هو 20 – 25= – 5، وأما الفرق بين ما هو متوقع وما حدث فعلاً فيما يتصل بالصورة فهو 30 – 25=+5، فإذا قمنا بتربيع كل من هذين الفرقين، فإننا نحصل على $(-5)^2 = 25$ و$(+5)^2 = 25$

ولتطبيق المعادلة السابقة لابد من قسمة كل من هذين الفرقين على التكرار المتوقع أي:

$$1 = \frac{25}{25}، 1 = \frac{25}{25}$$

وتكون قيمة كا^2 بالتالي 1+1 =2

ويمكن تطبيق المعادلة التالية على المثال السابق، وهي صيغة مختصرة للصيغة السابقة:

$$كا^2 = \frac{(ت_1 - ت_2)^2}{ت_1 + ت_2}$$

حيث يشير الرمز $ت_1$ إلى التكرار الأكبر

و $ت_2$ إلى التكرار الأصغر.

وبالتعويض عن قيم المثال السابق نحصل على:

$$\frac{(30 - 20)^2}{30 + 20} = 2$$

وهي النتيجة نفسها التي حصلنا عليها من المثال السابق.

ولحساب دلالة $كا^2$ لابد من حساب درجات الحرية وهي في هذا المثال $2 - 1 = 1$ (حيث يشير الرقم 2 إلى عدد فئات التكرار والرقم 1 إلى عدد القيود الإحصائية). وبالنظر في الجداول الإحصائية الخاصة بـ $كا^2$ عند درجة الحرية 1 نجد أن القيمة المتحصلة لـ $كا^2$ وقدرها 2 اقل من قيم $كا^2$ الموجودة في الجداول عند مستويي الثقة 95٪ و99٪ مما يشير إلى أن الفرق غير دال إحصائياً عند أي من هذين المستويين.

إن المثال السابق يتيح ظهور احتمالين فقط (أحدهما للصورة وهو بنسبة 50٪ أو ½ والثاني للكتابة وهو أيضاً بنسبة 50٪ أو ½) ولكن هناك حالات معينة تتيح ظهور ثلاثة احتمالات أو أكثر، وهنا لابد من قسمة مربع الفرق بين التكرار الواقعي والتكرار التجريبي على 3 بدلاً من 2 (وعلى 4 أو 5 أو أكثر إذا كانت الاحتمالات 4 او 5 او اكثر).

فلو افترضنا أن أحدهم طبق استطلاعاً للرأي على مجموعة من المفحوصين، وكانت بدائل الإجابة ثلاثة بدائل أو احتمالات وهي نعم، لا، لا أدري فيمكن أن نتوقع ظهور أي من هذه الاحتمالات بنسبة (1) إلى (3) من العدد الكلي للإجابات على كل من البنود التي يضمها الاستطلاع، مما يستدعي قسمة مربع الفرق بين التكرار الواقعي والتكرار المتوقع على 3. ويظهر ذلك في المثال الذي يوضحه الجدول التالي (38):

بدائل الإجابة	نعم	لا	لا أدري	مج
التكرار	120	80	40	240

إن التكرار المتوقع لكل من البدائل الثلاثة هو:

$$80 = \frac{240}{3}$$

وبالتعويض عن القيم في المعادلة:

$$\text{كا}^2 = \text{مج} \frac{(\text{ت و} - \text{ت م})^2}{\text{ت م}}$$

نحصل على:

$$\text{كا}^2 = \frac{(80-120)^2}{80} + \frac{(80-80)^2}{80} + \frac{(80-40)^2}{80}$$

$$= 20 + 0 + 20 = 40$$

ولتحديد ما إذا كنت قيمة كا^2 المتحصلة في المثال السابق (والمعبرة عن الفروق بين تكرارات الإجابات عن أحد البنود) دالة إحصائياً لابد من حساب درجات الحرية وهي في هذه الحالة 3 – 1=2 (حيث يشير الرقم 3 إلى عدد فئات

الإجابة و1 إلى عدد القيود). وبالرجوع إلى الجداول الإحصائية الخاصة بدلالة كا^2 عند درجة الحرية 2 نجد أن قيمة كا^2 التي حصلنا عليها أعلى من القيم الموجودة في الجدول عند مستوى الدلالة 5٪ و1٪ مما يشير إلى أن الفروق بين تكرارات الإجابات دالة إحصائياً بنسبة 99٪ من الثقة.

وثمة طرائق أخرى عديدة لحساب كا^2 نختار منها الطريقة الخاصة بحساب كا2 من الجدول الرباعي والتي يوضحها المثال التالي:

طبق اختبار في القدرة العددية على مجموعة من البنين عددهم 50 ومجموعة من البنات عددهن 40 وكان عدد من أعطى إجابات صحيحة على أحد البنود من البنين هو 30 ومن البنات هو 12، وذلك كما في الجدول التالي:

الجدول (39): حساب كا^2 من الجدول الرباعي:

الجنس	رسوب	نجاح	مج
بنين	أ 20	ب 30	50
بنات	ج 28	د 12	40
مج	48	42	90

إن التكرار المتوقع للخلية (أ) في هذا الجدول هو

$$26.66 = \frac{48 \times 50}{90}$$

وبتطبيق الصيغة العامة لحساب كا^2 على هذه الخلية تكون قيمة كا^2 لهذه الخلية هي:

$$1.66 = \frac{^2(26.666 - 20)}{26.666}$$

وأما التكرار المتوقع للخلية (ب) فهو:

$$23.33 = \frac{42 \times 50}{90}$$

وبتطبيق الصيغة العامة لحساب كا^2 على هذه الخلية تكون قيمة كا^2 للخلية (ب) هي:

$$1.90 = \frac{^2(23.33 - 30)}{23.33}$$

وأما التكرار المتوقع للخلية (ج) فهو:

$$21.33 = \frac{48 \times 40}{90}$$

وبتطبيق الصيغة العامة لحساب كا^2 تكون قيمة كا^2 للخلية (ج)هي:

$$2.08 = \frac{^2(21.33 - 28)}{21.33}$$

وأما التكرار المتوقع للخلية (د) فهو:

$$18.66 = \frac{42 \times 40}{90}$$

وقيمة كا^2 لهذه الخلية هي:

$$2.37 = \frac{(18.66 - 12)^2}{18.66}$$

والقيمة الكلية لـ كا^2 هي مجموع قيم كا^2 للخلايا الأربع أي:

$$1.66 + 1.90 + 2.08 + 2.37 = 8.01$$

وبما أن درجات الحرية = عدد الأعمدة $- 1 \times$ عدد الصفوف $- 1$

فإن درجات الحرية في الجدول الرباعي هي:

$$(2 - 1) \times (2 - 1) = 1$$

وبالرجوع إلى جداول الدلالة الإحصائية الخاصة بـ كا^2 لدرجة الحرية 1، يتبين أن القيمة المتحصلة وهي 8.01 أعلى من القيم الموجودة في هذا الجدول لمستوى الدلالة 5٪ و1٪، مما يدل على أن الفرق المحسوب بين البنين والبنات دال إحصائياً بنسبة 99٪ من الثقة.

المراجع

أولاً: المراجع العربية:

- أبو النيل، محمود السيد (1987): الإحصاء النفسي والاجتماعي والتربوي. دار النهضة بيروت.
- أبو حطب، فؤاد، وسيد أحمد عثمان (1979أ): التقويم النفسي. مكتبة الأنجلو المصرية، القاهرة.
- أبو لبدة، سبع محمد (1979): مبادئ القياس النفسي والتقييم التربوي للطالب الجامعي والمعلم العربي. جمعية عمال المطابع التعاونية، عمان.
- أحمد، محمد عبد السلام (1960): القياس النفسي والتربوي (ج1). مكتبة النهضة المصرية، القاهرة.
- البهي السيد، فؤاد (1978): علم النفس الإحصائي وقياس العقل البشري. دار الفكر العربي، القاهرة.
- البهي السيد، فؤاد (بلا تاريخ): الجداول الإحصائية لعلم النفس والعلوم الإنسانية الأخرى. دار الفكر العربي، القاهرة.
- الروسان، فاروق (1996): أساليب القياس والتشخيص في التربية الخاصة. دار الفكر للطباعة والنشر والتوزيع، عمان.
- تايلر، ليونا (ترجمة سعد جلال) (1975): الاختبارات والمقاييس النفسية. القاهرة.
- التير، مصطفى عمر (بلا تاريخ): مقدمة في مبادئ وأسس البحث الاجتماعي. المنشأة الشعبية للنشر والتوزيع والإعلان، الجماهيرية العربية الليبية.
- ثورندايك، روبرت، واليزابيث هيجن (ترجمة عبد الله زيد الكيلاني وعبد الرحمن عدس) (1989): القياس والتقويم في علم النفس والتربية. مركز الكتب الأردني، عمان.
- جابر، جابر عبد الحميد وأحمد خيري كاظم: مناهج البحث في التربية وعلم النفس. القاهرة، دار النهضة العربية، الطبعة الثانية، 1987.

- جلال، سعد (2001): القياس النفسي (المقاييس والاختبارات). دار الفكر العربي، القاهرة.
- حسن، عبد الباسط محمد: أصول البحث الاجتماعي: القاهرة، مكتبة وهبة، الطبعة الثامنة، 1982.
- حسن، عبد الباسط محمد: أصول البحث الاجتماعي: القاهرة، مكتبة وهبة، الطبعة الثامنة، 1982.
- حليمي، عبد القادر (1985): مدخل إلى الإحصاء. منشورات عويدات، بيروت- باريس.
- حمدان، محمد زياد (1986): تقييم التحصيل. دار التربية الحديثة، عمان.
- حمدان، محمد زياد: البحث العلمي كنظام، دار التربة الحديثة، عمان 1989.
- عبد الرحمن، سعد (1998): القياس النفسي (النظرية والتطبيق). دار الفكر العربي، القاهرة، ط3.
- علام، صلاح الدين محمود (2000): القياس والتقويم التربوي والنفسي (أساسياته وتطبيقاته وتوجّهاته المعاصرة). دار الفكر العربي، القاهرة ،ط1.
- السيد، محمود أحمد وامطانيوس مخائيل (1989): نماذج من الاختبارات الموضوعية في اللغة العربية للمرحلة الإعدادية (المتوسطة).المنظمة العربية للتربية والثقافة والعلوم، تونس.
- الغريب، رمزية (1970): التقويم والقياس النفسي والتربوي. مكتبة الأنجلو المصرية، القاهرة.
- عدس، عبد الرحمن (1989): دليل المعلم في بناء الاختبارات التحصيلية. إدارة البحوث التربوية، تونس.
- عودة، أحمد سليمان (1985): القياس والتقويم في العملية التدريسية. جامعة اليرموك.
- عوض. عباس محمود (1990): القياس النفسي بين النظرية والتطبيق. دار المعرفة الجامعية، الإسكندرية.

- فرج، صفوت (1980): القياس النفسي. دار الفكر العربي، القاهرة.
- لندفل، س.م (ترجمة عبد الملك الناشف وسعيد التل) (1968): أساليب الاختبار والتقويم في التربية والتعليم. المؤسسة الوطنية للطباعة والنشر، بيروت.
- مادوس ،جورج ف وبنجامين س بلوم و ج.توماس هاستنجس (ترجمة محمد أمين المفتي وآخرين) (1983): تقويم تعلم الطالب التجميعي والتكويني. دار ماكجروهيل للنشر، الرياض.
- موسى، فاروق عبد الفتاح علي (1990): القياس النفسي والتربوي للأسوياء والمعوقين. مكتبة النهضة المصرية، القاهرة.

ثانياً: المراجع الروسية:

- بوداليف، أ.أ.، وستولين ،ب.ب. (1987): التشخيص النفسي العام. إصدار جامعة موسكو.
- سيفكينا، م.أ.: الاستبانة ومكانتها ضمن أدوات البحث التربوي. مجلّة التربية السوفييتية، العدد 9، 1986.
- سوخودولسكي (1972): أسس الإحصاء الرياضي للعاملين في علم النفس.دار التقدم، موسكو.
- كابانوفا، ل.ر.: مشكلات البحث في التربية وعلم النفس، إصدار دار الكتاب، موسكو، 1982.
- ليرنر، م.أ.: تطور البحث التربوي، دار العلم، موسكو، 1983.
- ماتيوشكين، ج.د.: التعليم المشكّل ومسائل البحث التربوي، إصدار دار التقدّم، موسكو، 1981
- تاليزنيا، ن. ف. (1978):التحكم بعملية استيعاب المعارف. إصدار جامعة موسكو، موسكو.
- تاليزنيا، ن. ف. (1980): مبادئ علم النفس السوفيتي ومسائل تشخيص النشاط المعرفي. في كتاب " التشخيص النفسي والمدرسة"، تالين.

- تاليزينا، ن. ف.(1986): مقدمة كتاب "التشخيص النفسي" لمؤلفه ج.فيتيسلاك. موسكو، دار التقدّم.
- فيتيسلاك، ج. (1986): التشخيص النفسي. (نقله من الألمانية إلى الروسية: بوتكين، أ وزميله). موسكو، دار التقدّم.
- لولر ،ج. (1982): حاصل الذكاء، الوراثة والعنصرية. موسكو، دار التقدّم.
- يرماكوفا، س.ي.: مناهج البحث التربوي، دارالعلم، موسكو،1979 .

ثالثاً: المراجع الإنكليزية:

- Ahmann, J.S. &Glock, M.D.(1975).Evaluating Pupil Growth. Boston: Allyn and Bacon ،Inc.
- Aiken, L.R. (1996). Assessment of intellectual functioning
 (2nd ed.). New York: Plenum.
- Anastasi, Anne. (1982).Psychological Testing (Fifth Edition). Macmillan Publishing Co.Inc. New York.
- Anastasi,A.& Urbina, S. (1997). Psychological Testing(7th
 ed.). Prentice-Hall, Inc.
- Atkinson, J. W.,& Raynor, J. O. (Eds.).(1974). Motivation and achievement. Washington, DC:Winston.
- Bloom,B.S.(1968). learning for mastery. UCLA-CSEIP evaluation comment. 1.
- Brown, F.G.(1983). Principles of educational and psychological testing.new York: Holt-Rinehart and Winston.
- Butler,S.M. &Mcmum,n.D.(2006), A Teatcher's guide to glassroom assessment.John wiley & Sons.
- Cohen,R.J and Swerdlik, M. E.(2005), Psychological testing and Assessment: An Introduction to tests and Measurement.,5th edition,Boston,McGraw-Hill.

- Dahlstrom. W. G. (1993). Tests: Small samples ،large consequences. American Psychologist, 48, 393-399.
- Glaser, R.A.(1971). Criterion-referenced test. In criterion-referenced measurement (J.W. popham, ed.).Englewood cliffs. educational technology publications,.
- Gronlund, N. E.(1971). Measurement and Evaluation In Teaching (2.nd ed.). New York: Macmillan.
- Gronlund, N. E.(1977). Constructing achievement tests (2.nd ed.). Englewood cliffs, N.J.: prentice-Hall.
- Gronlund, N. E.(1978). Stating behavioral objectives for classroom instruction(2.nd ed.). New York: Macmillan.
- Hilgard, E. R.(1989). The early years of intelligence measurement. In R. Linn(Ed.) ،Intelligence Measure-ment ،theory ،and public policy. Urbana, IL:University of Illinois Press.
- Hoffman, B.(1962).The tyranny of testing. New York: Crowell-Collier.
- Hopkins,C. and Antes,R.(1985) Classroom measurement and evaluation (2.nd ed.),Itasca, Illinois, F.E.Peacock publishers,Inc.
- Matarazzo,J.D.(1992). Psychological testing and assessment in the 21st century. American Psychologist, 47,1007-1018.
- Mehrens W.A. & Lehmann, I. J.(1973). Measurement and evaluation in education and psychology ،Holt ،Rinehart & Winston, Inc. New York.
- Murphy,K.R.&Davidshofer,C.O.(2001).Psychological testing,Principles and applications,Prentice Hall.
- Nichols&Mittlholtz.(1996).Constructing the Concept of Aptitude, San Diego,Academic press.
- Nunnally,J.C.(1982). Psychometric theory (2nd ed.). New york:McGraw-Hill.

- Osterlind,S.J.(2006). Modern Measurment: Theory، Principles ،and Applications of Mental Appraisal.Merrill Prentice Hall.u
- Popham,W.J.(1981). Modern Educational measurement. Englewood Cliffs, N.J.prentice-Hall.
- Sternberg, R. J., & Detterman ،D. K. (Eds.). (1986). What is intelligence? Contemporary viewpoints on its nature and definitions. Norwood, NJ: Ablex.
- Stufflebeam, D.L. (1971). Educational evaluation and decision making.Itasca, Illinois: F.E.Peacock publishers.
- Tuckman.B. W.(1975).Measuring educational Outcomes. Harcourt Brace Inc. New York. Jovanovich.

ثبت المصطلحات العلمية

A

Achievement	التحصيل (أو الإنجاز)
Achievement age	العمر التحصيلي
Achievement subtests	الاختبارات الفرعية التحصيلية
Achievement test	الاختبار التحصيلي
Age deviation scores	درجات أعمار انحرافية
Age equivalents	المكافئات العمرية
Age norms	معايير العمر (المعايير العمرية)
Aggressiveness	السلوك العدواني
Alpha coefficient	معامل ألفا
Analysis of variance	تحليل التباين
Assessment	تقدير
Associational Fluency	طلاقة التداعي
Availability	جدوى (المحك) أو مدى تيسره
Average deviation	الانحراف المتوسط

B

Basal age	العمر الأساسي (القاعدي)
Basal level	المستوى القاعدي أو الأساسي
Blueprint (table of specification)	جدول المواصفات

C

Category	الزمرة (الفئة)
Ceiling age	السقف العمري (العمر الأقصى)
Ceiling level	المستوى الأقصى او السقف
Central tendency	النزعة المركزية
Check list	قائمة الرصد
Chronological age	العمر الزمني
Classroom test	الاختبار الصفي
Coefficient of stability	معامل الاستقرار

Combination aptitude achievement tests	اختبارات التحصيل والاستعداد
Concurrent validity	الصدق التلازمي (الملازم)
Consistency	اتّساق
Construct validity	الصدق البنيوي (الافتراضي)
Constructs	السمات او التكوينات الفرضية
Content	المحتوى
Content Referenced test	اختبار مرجعي المحتوى (اختبار بدلالة المحتوى)
Content validity	صدق المحتوى
Continuum	اتصال (متصل)
Contrasted groups	الفرق المتقابلة (الجماعات المتعارضة)
Correlation coefficient	معامل الارتباط
Convergent validation	الصدق التقاربي للاختبار
Criterion	المحكّ
Criterion group	الجماعة المحكية
Criterion referenced test	الاختبار المحكي المرجع
Criterion referenced validity	الصدق المحكّي (الصدق بدلالة محك)
Cross-Cultural tests	الاختبارات عبر الثقافية (العالمية)
Cutoff scores	درجات القطع

D

Deviation I.Q	حاصل الذكاء الانحرافي
Decile	العشير
Derived standard scores	الدرجات المعيارية المعدّلة (المشتقة)
Diagnostic subtests	الاختبارات الفرعية التشخيصية
Diagnostic test	الاختبار التشخيصي
Difficulty index	معامل التمييز
Difficulty range	نطاق أو مدى الصعوبة
Discriminant validation	الصدق التمييزي (أو التباعدي) للاختبار

Discrimination index	معامل التمييز
Discrimination power	القدرة التمييزية
Dispersion	التشتت
Distractors	المشتتات (المموهات في أسئلة الاختيار من متعدد)
Distribution of scores	توزيع الدرجات
E	
Equation Building	بناء المعادلات
Equivalent forms	الأشكال المتعادلة
Equivalent forms of a test	الأشكال المتعادلة (طريقة الأشكال المتعادلة في حساب الثبات)
Error of measurement	خطأ القياس
Error variance	التباين الناتج عن الخطأ (او تباين الخطأ)
Exploratory	الاستكشافي
Extended age range	المدى العمري الموسّع
F	
Face validity	الصدق الظاهري (الظاهر)
Factor analysis	التحليل العاملي
Factorial validity	الصدق العاملي
Frequency curve	المنحني التكراري
Frequency distribution	التوزيع التكراري
Frequency polygon	المضلع التكراري
G	
Grade	الدرجة
G.Score	الدرجة الجيمية
General factor	العامل العام
Grade equivalents	المعادلات الصفية
Grade norms	معايير الصف (المعايير الصفية)

Grade-based instead of age-based percentiles	درجات مئينية ترتكز على الأداء الصفي بدلاً من الأداء العمري
Group factor	العامل الطائفي
Group test	الاختبار الجمعي
Guessing correction	التصحيح من اثر التخمين

H

Histogram	المدرج التكراري
Homogeneity coefficient	معامل التجانس

I

Incidental sample	العينة العرضية
Index of validity	مؤشر الصدق
Individual test	الاختبار الفردي
Individualized	التفريدي
Intellectual efficiency	الكفاية العقلية
Intelligence quotient	حاصل الذكاء (أو نسبته)
Intelligence test	اختبار الذكاء
Interindividual	بين الأفراد
Interitem consistency	الاتساق الداخلي بين بنود الاختبار
Internal consistency	الاتساق الداخلي
Item	البند (الفقرة الاختبارية)
Item analysis	تحليل البنود (الفقرات)

K

Kuder-Richardson reliability	الثبات بطريقة كودر – ريتشاردسون

L

Learning age	العمر التعليمي
Learning quotient	حاصل التعلّم
Logical validity	الصدق المنطقي

M

Manual	دليل (الاختبار)
marks	العلامات
Mean	المتوسط
Measure	مقياس
Measurement	قياس
Measurement error	خطأ القياس
Median	الوسيط
Mental age (mA)	العمر العقلي
Mental Scale	المقياس العقلي
Motivation	الدافعية
multiple - choice items	بنود (أسئلة) الاختيار المتعدد
Norm	المعيار
Norm referenced test	الاختبار المعياري المرجع
Normal curve	المنحني الطبيعي (السوي)
Normal distribution	التوزيع السوي (الاعتدالي)
Normalized standard scores	درجات معيارية اعتدالية
Normative or standardization sample	عينة التعيير (أو عينة التقنين)

O

One-factor model	النموذج العاملي الأحادي
Operational definition	التعريف الإجرائي
Overall composite score	الدرجة الكلية الإجمالية

P

Parallel forms	الأشكال المتعادلة
Percentile	الميئينات
Percentile band	الزمرة المئينية
Percentile norms	المعايير الميئينية
Percentile rank	الرتبة المئينية

Performance tests	الاختبارات الأدائية (غير اللفظية)
Pilot study	الدراسة الاستطلاعية
Placement	تحديد المسار التعليمي(الوضع في المكان المناسب)
Placement decisions	قرارات المسار (أو الوضع في المكان المناسب)
Population	المجتمع الأصلي
Power test	اختبار قوة
Predictive validity	الصدق التنبؤي
Pretest	اختبار قبلي
Principal components	المكوّنات الأساسية
Profile	الصفحة النفسية (البروفيل)
Prognostic tests	الاختبارات التنبؤية
Purposive sample	العينة المقصودة

Q

Qualitive	كيفي
Quantative	كمي
Quartile	الرُبيع
Questionnaire	الاستبانة (الاستخبار)

R

Random error	الخطأ العشوائي
Random sample	العينة العشوائية
Range	المدى
Rank	رتبة (مرتبة)
Rater reliability	ثبات المقدّر
Rating reliability	ثبات التقدير
Raw score	العلامة الخام
Reliability	الثبات (الموثوقية)
Reliability coefficient	معامل الثبات

S

Sample	عينة
Sampling error	خطأ العينة (خطأ المعاينة)
Score patterns	انماط الدرجات
Scorer reliability	ثبات المصححين
Scoring	عملية وضع الدرجات (تصحيح الاختبار)
Scoring key	مفتاح التصحيح
Self-report	التقرير الذاتي
Skewed distribution	التوزيع الملتوي
Social age	العمر الاجتماعي
Social quotient	"معيار" النسبة الاجتماعية
Socioeconomic status	مؤشرات الحالة او المركز الاجتماعي و الاقتصادي
Specific factor	عامل خاص (نوعي)
Split-half method	طريقة التنصيف
Split-halves	التنصيف (التجزئة إلى نصفين)
Standard age scores (SAS)	درجات معيارية عمرية
Standard deviation	الانحراف المعياري
Standard error	الخطأ المعياري
Standard error of measurement (SEM)	الخطأ المعياري للقياس
Standard score	العلامة المعيارية (الدرجة المعيارية)
Standard Score Norms	معايير الدرجة المعيارية
Standardization	عملية التعيير (أو المعايرة او التقنين)
Standardized test	الاختبار (الرائز) المقنّن أو المعيّر
Stanine	التسيع (معيار التسيع)
Stanine score	علامة ستانين (معيار التسيع)
Statistical table	الجدول الاحصائي
Stratified sample	العينة الطبقية
Supplementary norms	معايير إضافية

T

T.Score	الدرجة التائية
Table of specifications	جدول المواصفات
Task	المهمة
Teacher-made test	اختبار المعلم
Test	الاختبار (الرائز)
Test administration	إجراء الاختبار
Test ages	أعمار اختبارية
Test battery	بطارية (طاقم) الاختبارات
Test items	البنود الاختبارية (الفقرات)
Test manual	دليل الاختبار
Test taking experience	خبرة أخذ الاختبار
Test-retest reliability	الثبات بطريقة الإعادة
Total standard score (TSS)	الدرجة المعيارية الكلية
Transformations-T	التحويلات
True score	الدرجة الحقيقية

U

Unidimensional	أحادي البعد

V

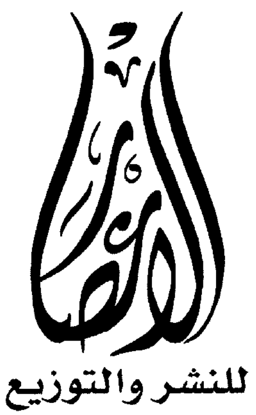
Validation process	عملية التحقق من الصدق (تأسيس الصدق)
Variance	التباين
Variety	التنوّع

Z

Z-Score	العلامة المعيارية (الذالية)


للنشر


MOHAMED KHATAB


الأصالة
للنشر والتوزيع

بناء الاختبارات والمقاييس النفسية والتربوية وتقنيتها









الأردن - عمان - وسط البلد - ش. الملك حسين - مجمع الفحيص التجاري
هاتف : 96264646208+ فاكس : 96264646470+

الأردن - عمان - مرج الحمام - شارع الكنيسة - مقابل كلية القدس
هاتف : 96265713906+ فاكس : 96265713907+
جوال : 00962-797896091

info@al-esar.com - www.al-esar.com

دار الاعصار العلمي